ASPECTS OF MODELLING IN MEDICINE

A thesis

presented for the degree of

Doctor of Philosophy in Electrical Engineering

in the

University of Canterbury,

Christchurch, New Zealand

by

P.B. HEFFERNAN B.E. (HONS)

1981

ABSTRACT

Three applications of models in medicine are considered.

A computer-aided learning program for teaching the dynamics of uptake and distribution of the inhalational anaesthetic halothane is described. The program is based on a compartmental model which simulates the action of halothane on ventilation and on the cardiovascular system. The program presents the model to the student in four forms: one with no changes in circulation or respiration, one with the cardiovascular effects of halothane included, one with respiratory effects only, and one with both of these effects combined. The student can study the importance of halothane's influence on respiration and blood circulation by comparing results from simulations on different models. The simulation is presented as graphs continuously displayed on an alphanumeric visual display terminal. Interaction with the program is possible at all times to change the simulation speed, the variables being graphed, the halothane fraction, and the fresh gas flow. A pilot evaluation of the program as it is used during a tutorial shows a highly significant improvement in the students' answers after the tutorial using a 'before and after' questionnaire. The students showed an understanding of the program's display and model limitations. This encourages the further use of the program.

The images obtained by Computed Tomography are degraded if there is any body motion while the projections are being measured. This blurring is assessed by considering motions of individual points within body cross-sections. An effective radius of body movement is defined and a statistical description of the blurring is developed. Both continuous and discrete movements are investigated. It is found that the resolution limit set by the blurring is almost always twice the effective radius, although the blurring can sometimes be negligible if the movement occurs spasmodically during the measurement of the projections.

In some situations of technical and scientific interest, only small numbers of projections are available. For these, conventional reconstruction techniques produce unacceptable images. Several methods for increasing the number of projections by interpolating between given projections are introduced and compared. It is found that schemes which

employ *a priori* information give images which are superior to those which do not utilise such information. The general structure, rather than the specific details (e.g. whether or not the underlying functional is of *maximum entropy* type) of an iterative scheme determines its usefulness as an interpolation method. It is shown how the convergence rates of such iterative schemes can be accelerated.

ACKNOWLEDGEMENTS

I am especially grateful to my joint supervisors, Professor R.H.T. Bates and Dr A.E. McKinnon. I thank them for their patient reviewing of the many drafts of this thesis.

I would like to thank Professor J.M. Gibbs of the Department of Anaesthesia, Christchurch Clinical School of Medicine, for his considerable help with the work presented in Chapter 2. I would also like to thank the following for their help on that project: Dr C.G. Male and Dr F.M. Davis of the Christchurch Clinical School of Medicine for their advice on the medical aspects of halothane anaesthesia and for their enthusiastic use of the program for teaching; Mr A.W. Stewart for his help with the statistics; Professor W.W. Mapleson of the Department of Anaesthetics, Welsh National School of Medicine, Cardiff, Wales, whose careful refereeing of the two papers presented to "Anaesthesia" resulted in several modifications to the work presented in Chapter 2.

I would like to thank my colleagues from the Electrical Engineering Department and all those who contributed to the seminars of the Biomedical Engineering Group. In particular I would like to thank Dr G.C. McKinnon whose suggestions led to part of the work which is presented in Chapter 4.

During the course of the work reported herein I was the grateful recipient of a University Grants Committee Postgraduate Scholarship. The partial financial support of Technicare Inc. for the computational studies reported in Chapters 3 and 4 is also acknowledged.

PREFACE

The work reported in this thesis started in 1978 when I enrolled for a Ph.D. in Electrical Engineering under the joint supervision of Professor R.H.T. Bates of the Electrical Engineering Department, University of Canterbury, and Dr A.E. McKinnon, who was then Director of the Computer Laboratory, Christchurch Clinical School of Medicine. I initially intended to look at some aspects of Computed Tomography, but as an introduction to research work, and with the cooperation of Professor J.M. Gibbs of the Department of Anaesthetics, Christchurch Clinical School of Medicine, I embarked on a small project to design a computer aided learning program to aid the teaching of the uptake and distribution of the inhalational anaesthetic halothane.

The aspects of Computed Tomography (CT) and computer aided learning studied in this thesis are tied together by the concept of modelling. Chapter 1 explains this connection and introduces some basic ideas and techniques which are used in the remainder of the thesis. A brief overview of computer aided learning is given in Chapter 1 and the basic theory of computed tomography is presented. There can be loss of image quality if any one of the basic premises of CT is not kept. Chapter 1 concludes with a review of such mechanisms and briefly details the approaches taken to reduce their effects and to characterise the image degradation.

The second chapter describes the program I constructed for Computer Aided Learning (CAL). The program simulates the uptake and distribution of halothane and is designed as an interactive tutorial aid. The development of such programs suffers from the difficulties associated both with the derivation of a model and with the design of the computer/student interface. I derived this original model from the data available in the literature. The clinical experience of members of the staff of the Department of Anaesthetics was used to overcome the difficulties caused by insufficient and inconsistent data. The resulting compartmental model forms the basis of the CAL program. This program has been used routinely for teaching fifth year medical students. The members of the staff who use it find the program easy to use and a successful teaching aid. In order to verify its usefulness, I carried out a pilot evaluation study which is also reported in Chapter 2.

Chapter 3 investigates an application of modelling in computed tomography. Computed tomography is a useful diagnostic X-ray technique which produces cross-sections of the body. Series of projections are taken around the body through the plane of interest. These projections are used to form the image of the cross-section. Current CT scanners can measure a full set of projections in less than 2 seconds. However, even during such short scan times, there may be significant patient movement. For instance, there could be involuntary muscle movement, breathing motion, peristalsis, or cardiovascular movement. Any such movement causes blurring in a reconstructed image. In practice, the details of the movement are rarely known, so that it is not possible to correct for it. The aim of the analysis presented in Chapter 3 is to characterise the blurring due to motion. The approach I adopt is based on models of the types of movement that can be expected in practice. The advantage of this is that it allows the blurring due to classes of movement to be characterised analytically. Thus I obtain quite general results without resorting to prolonged computational experiments.

The work reported in Chapter 4 is concerned with an aspect of computed tomography which was originally suggested by some of the work of Graeme McKinnon while he was a graduate student at this University. He developed a technique for imaging the beating heart with a conventional CT scanner. However his technique produces only small numbers of projections at each phase of the heart cycle. Conventional reconstruction techniques do not make the best use of the available measurements in such situations. Thus I have examined several reconstruction techniques in an effort to obtain better images. The methods presented in Chapter 4 are based on a model for the image. This model embodies whatever *a priori* information is available. For example, it is usually known that the image is real and non-negative, and an estimate of its size is often available. Maximum entropy and other statistically inspired measures also contribute *a priori* information. In Chapter 4 I compare eight reconstruction methods, six of which employ a model for the image. The more successful of the schemes are derived by posing the reconstruction problem as an optimisation problem. This sort of approach is not new, but this new application to the problem caused by reconstruction when there are few measured projections has allowed me to develop a reconstruction framework into which a variety of image models can be incorporated. The four methods based on this approach give the best reconstructions and appear to be worthy of further study and

implementation with measured data.

The thesis concludes with Chapter 5, which presents conclusions and suggestions for further research.

During the course of the work presented in this thesis the following presentations and papers have been prepared:

Heffernan P.B., Gibbs J.M., and McKinnon A.E. 1979. "A Simulation of Anaesthetic Uptake and Distribution", Presented at the 1978 meeting of the Anaesthetic Research Group of Australia and New Zealand, Abstract in Anaesthesia and Intensive Care VII, 198.

Bates R.H.T. and Heffernan P.B. 1980. "Image reconstruction from projections. V: Blurring due to object movement", Optik 56, 101-112.

Heffernan P.B., Gibbs J.M., and McKinnon A.E. 1980. "Teaching halothane uptake and distribution by computer simulation", Presented at the Conference of the Australasian College of Physical Sciences and Engineering in Medicine and Biology, Christchurch, New Zealand, August 1980 (proceedings page 25).

Heffernan P.B. and Bates R.H.T. 1981. "Image reconstruction from projections. VI: Comparison of interpolation methods", Optik (in press).

Heffernan P.B., Gibbs J.M., and McKinnon A.E. 1982. "Teaching halothane uptake and distribution. I: A computer simulation program", Anaesthesia (in press).

Heffernan P.B., Gibbs J.M., and McKinnon A.E. 1982. "Teaching halothane uptake and distribution. II: Evaluation of a computer simulation program", Anaesthesia (in press).

TABLE OF CONTENTS

1. INTRODUCTION

1.1 OVER-VIEW

Three applications of modelling in medicine are considered in this thesis. The first uses a mathematical model for the uptake and distribution of an inhalational anaesthetic. This is implemented on a computer as a teaching aid. The other two applications are in a particular field of patient imaging - Computer Aided Tomography.

The fundamental concepts of system and model are introduced in §1.2 and §1.3. §1.4 discusses the advantages and disadvantages of Computer Aided Learning (CAL) and presents a classification scheme for CAL programs. An area where modelling can be useful for teaching is also investigated in §1.4. The principles of Computer Aided Tomography, nowadays called Computed Tomography (CT), are explained in §1.5. The kinds of image degradation that commonly occur in actual practice are discussed in §1.6. The chapter concludes in §1.6 with a review of approaches to artefact characterisation and correction.

1.2 SYSTEM

The concept of a system has appeared in fields as diverse as biology and economics. A review of the many different types of systems is given by McKinnon (1973). A general definition of a *system*, taken from control theory (Elgerd, 1967, p. 1), is a "collection of objects united by some form of interaction or interdependence". However, despite this definition, any discussion or analysis of a real system implies some conceptual representation of it. Thus the notions of system and model necessarily overlap. This point is not considered further here since it is convenient to use the above definition of a system isolated from any implicit modelling.

Systems can often be described as hierarchies of *sub-systems*. Consider, for example, the administration of some inhalational agent to a patient in an operating theatre. The system could be considered to contain the anaesthetist, patient, vapouriser, delivery circuit and surgeon (Bagshaw, 1980). The anaesthetist assesses the patient's state by visual, tactile and auditory feedback. One sub-system monitored by

the anaesthetist controls blood pressure: it has neural, endocrine and renal divisions. Since systems and sub-systems can also be defined in terms of their mechanisms, in the above example the neural division could be considered to be separate from the other divisions which control the blood pressure. This does not mean that they are functionally isolated. The effect of one part of a system on another can be considered to be an input (or control force) for the second part.

In control theory it is usual to describe a system in terms of controlling forces (inputs), system states and system outputs. A general system has many inputs and many outputs. A strict definition of the *state* of a system is the "minimum set of numbers or variables, the state variables, which contain sufficient information about the past history of a system to permit us to compute all future states of the system - assuming, of course, that all future inputs (*control forces*) are known and also the equations (bonds of interactions) describing the system" (Elgerd, 1967, p. 34). This definition leads to the important dual concepts of controllability and observability. A system is said to be *controllable* if it is possible to find a set of control forces which can transfer the system between two arbitrarily specified states in a specified finite time (Elgerd, 1967, p. 87). No real system is entirely controllable. In the above anaesthetic delivery example, the control forces could be the rate of administration of the drug and the oxygen delivery rate. The anaesthetist may desire to raise the brain level of the anaesthetic (the state) in a certain time. It is not trivial to decide whether or not this system is controllable in a given situation. A system is said to be *observable* if measurements of its outputs enable the system state to be completely identified (Elgerd, 1967, p. 104). Consider the blood pressure control sub-system in the above example. It is composed of several divisions, yet there is likely to be only one system output measurable: the blood pressure. The blood pressure control sub-system is unlikely to be observable from only this one measurement.

Systems are usefully classified as either linear or non-linear. *Linear* systems obey the principle of *superposition*, that is, the application of the sum of two inputs to a system produces an output which is equal to the sum of the two outputs which would be obtained if the inputs were applied separately (Elgerd, 1967, p. 46). Most of the systems that have been studied in the past are linear since the equations describing them are easier to solve than those describing

non-linear systems. However, many important systems are not even approximately linear. For example, consider the delivery of the inhalational anaesthetic halothane. An inspired halothane level gives some halothane level in a particular tissue. Doubling the inspired level does not generally cause the tissue level to be even approximately doubled. This is because halothane depresses ventilation and cardiac output (cf. Munson *et al.*, 1973) and alters the distribution of blood flow to the tissues (cf. Smith *et al.*, 1972). Thus the superposition principle cannot be applied, which shows that this system is non-linear.

Despite the example quoted above, non-linear systems are often linearised for small perturbations around a reference state. Thus it might be possible to study the anaesthetic delivery problem in terms of linear theory, but this approach would only be valid for a limited range of system states.

Systems are studied for three main reasons. The first is to analyse their behaviour so as to enable predictions to be made in different situations. This is the usual aim of studies in the medical area. The second reason for studying a system is to enable a control strategy to be developed. In the anaesthetic delivery example, an impracticably complete understanding of the system dynamics would be necessary to enable the design of a machine to completely replace the anaesthetist. There have been numerous attempts to design an automatic anaesthesia system (cf. Salamonsen, 1976; Suppan, 1977; or Mapleson, 1979). However, current knowledge is insufficient to successfully design such a machine. The machines that have been built still require an anaesthetist to monitor the patient and take corrective action when the machine is in error. The third reason for studying a system is to enable better systems to be designed and built.

1.3 MODELLING

A *model* is a representation of a system. This definition includes both physical and mathematical models. A summary of different types of models is given by McKinnon (1973). The models invoked in this thesis are expressed mathematically.

The important steps in model formulation are system analysis, model synthesis, parameter estimation, computer implementation, model testing, and prediction (McKinnon, 1973). This process is not

straightforward however. In general the modelling procedure is heuristic. It is usually necessary to compromise between accuracy and simplicity. Only for linear systems is there any well-defined modelling approach. Methods of parameter estimation for linear systems have recently been summarised by Bates (1981).

A model which closely matches a system could be described as *strong*. An example of this is a system which is modelled by a set of differential equations which completely define all system outputs for a given state and input. A *weak* model is one which only partly specifies a system. Consider, for example, the reconstruction of an image from a set of measurements. The assumption that the image is positive is an example of a weak model for the image.

In Chapter 2 a model is developed for the uptake and distribution of the inhalational anaesthetic halothane. The model is described by a set of differential equations which completely define its behaviour. In this sense it is a strong model. The model is used within a computer program which is designed as a teaching aid.

One of the assumptions of conventional Computed Tomography is that a stationary object is being imaged. If movement occurs while the projections are being measured, and no account of it is taken in the reconstruction process, the image will suffer from blurring and loss of resolution. The type of blurring due to a particular effect is known as an *artefact*. A general review of the sources of artefacts in Computed Tomography is given in §1.6. Chapter 3 investigates the artefacts due to different forms of object motion which could occur during a CT scan. The approach taken is to model these different forms of motion. A measure of the loss of resolution due to a given artefact is introduced in Chapter 3. The aim is to characterise object motion artefacts in Computed Tomography (CT).

Conventional CT machines routinely use more than 200 projections to produce an image (cf. McKinnon and Bates, 1981). However, as described in §1.6.2, there are some situations where only small sets of projections are measured. Chapter 4 investigates the image degradation caused by having only a small set of measured projections. Straightforward reconstruction gives poor images. However, extra projections can be obtained by interpolation to improve the image quality. Chapter 4 looks at several projection interpolation schemes. Methods based on weak models of the image are found to give the best reconstructed image from a limited

set of measured projections. In practice an image is formed in a computer as numbers on a grid which represents the image plane. The assumption that these image values are not only real but are also positive is known as *image positivity*. Image positivity is an example of the type of model which is examined in Chapter 4.

1.4 COMPUTER AIDED LEARNING

Many high hopes have been expressed during the past twenty years for the usefulness of computer aided learning (Hoffer *et al.*, 1975). However, computer aided learning (CAL) has not yet had the expected impact. Some of the blame for this must be laid with those proponents who pushed for computer-based rather than computer-aided instruction. They tried to create machines that would encompass all areas of teaching, not only those to which the computer is most suited.

Before continuing, it is convenient to categorise CAL programs into four classes of roughly ascending complexity. The first class groups 'question and answer' programs. These usually consist of *multiple-choice* questions (McIntyre, 1980). This form is well suited to a computer implementation as it is easy to assess the validity of answers. It is doubtful whether this type of program should strictly be classed as a learning program. It might make better sense to describe 'question and answer' programs as assessment tools. They are most useful for providing self-evaluation for the students, and for determining weaknesses in teaching schemes.

The second class of teaching program is known as the 'decision tree'. A branched-tree structure is the simplest format for a dynamic simulation. For medical teaching the common *patient management problems* fit into this class. An example is a program that teaches the management of chronic hypertension. At each stage the student has the choice of examining patient status or administering some drug. The computer responds with requested patient variables (and warns of emergency situations). This type of program has proved popular and useful in a teaching environment (Hoffer *et al.*, 1975).

A mathematical model of system behaviour characterises the third class of program. An example would be a program which graphs patient variables (e.g. heart rate and ventilation) as the student alters the

administration of some drug (e.g. programs in the 'Mac' series, Dickinson, 1977).

In the fourth class, which is an extension of the third class, the system model is comprehensive enough to allow skill learning. This means that the computer closely mimics the system which is being modelled. Flight trainers are examples of this class. With such machines pilots are taught to fly in an automated environment that simulates both the cockpit of the airplane and the flying conditions the pilot could face (cf. Chambers, 1977).

The above discussion highlights some of the advantages of CAL (particularly for medical teaching). One of the main advantages is that it provides individual teaching. The computer provides immediate feedback so that the student takes an active rather than passive role in the learning process. The computer is a tireless tutor which can adapt to the student's pace (cf. Lamb and Bates, 1978). CAL programs can also help to standardise teaching between institutions offering the same courses. These factors are common to all four of the classes of CAL described above. Apart from question and answer programs, the other types of CAL provide simulated patients to aid medical teaching. This has the following advantages:

(i) Simulated patients with a variety of illnesses are readily and consistently available.

(ii) The patient's progress can be simulated over any desired time span.

(iii) The student can assume full responsibility for patient care without jeopardizing the safety of a real patient.

(iv) The student can gain experience in situations (e.g. cardiac arrest) where actual practice would not be possible.

(v) The student can see the result of inappropriate decisions which would be hazardous with a real patient.

Chambers and Sprecher (1980) divide the difficulties with CAL into the following three categories (in order of importance):

(i) The need for teachers to move from established teaching aids which work and to try the new, and as yet unproved, methods.

(ii) The current disorganisation in CAL caused by the diversity of hardware and author languages.

(iii) The cost of hardware, existing programs, and trained staff.

Most teaching institutions have computers, but program development is expensive. CAL programs developed at one institution are not always easy to implement at another. The reason for this is the dependence of a CAL program on some particular facility (e.g. computer language, type of terminal, analogue/digital converter). This situation is not helped by the large number of languages available to a programmer. There are author languages such as Tutor (PLATO IV author group, 1974) which are suited to question and answer and branched-tree structure programs. A wide variety of languages, such as ISL (Braun *et al.*, 1977), is available for the design of simulation programs. However, there does not appear to be a language designed particularly for the development of CAL simulation programs. Generally, for a single programming effort it is expedient to resort to one of the general-purpose languages (e.g. FORTRAN) available on the particular computer.

There are several additional disadvantages of simulation-based programs that must be considered (Adkins and Pooch, 1977):

(i) A simulation model may become expensive in terms of manpower and computer time.

(ii) Extensive development time may be encountered.

(iii) Hidden critical assumptions may be present causing the model to diverge from reality.

(iv) Model parameters may be difficult to initialise. These may require extensive time in collection, analysis and interpretation of data.

CAL programs which employ mathematical simulations suffer from the difficulties associated both with design of interactive programs and with the development of a model. The program is based on a conceptual model of the system being studied. This conceptual model is then described in some language, such as the language of a computer, thereby creating a simulation program. This program is then debugged and tested for logical accuracy and, if necessary, further modified and refined. This activity is called *verification* of the simulation program (Chattergy and Pooch, 1977). Once the program has been verified to the programmer's satisfaction, it is then tested with input data from the real system. This activity is called the *validation* of the model.

Future developments in CAL should make use of an improved student/computer interface. Dialogue style, displays and graphics are areas of current research (cf. Miller and Thomas, 1977). The recent advent of

microcomputers with multisensory capabilities and low costs should result in significant increases in CAL usage at all levels of education. In addition to providing similar capabilities to minicomputers, however, some microcomputers also permit voice input and output, colour displays, high resolution graphics, and text editing (Chambers and Sprecher, 1980). Microcomputers also have the advantage that they are essentially portable. Their disadvantages are currently in the areas of processing speed and mass storage. Thus they are well suited to CAL programs which do not require large storage or processing speed and so are not suited to the implementation of sophisticated simulation programs. These limitations will become less important as faster and cheaper microcomputers with large storage capabilities become available. Even when CAL is administered by a large computer, microcomputers are proving useful as intelligent terminals (e.g. in the PLATO system, Smith, 1979).

Traditionally, learning in medicine has been something of an apprenticeship. In conjunction with skill-learning on the job, medical students take comprehensive courses in physiology, pharmacology, etc. Because most biological systems are both exceedingly complex and poorly understood, few mathematical simulations are available for medical teaching. Chapter 2 presents a particular application of CAL which is designed as an aid for teaching anaesthetic practice. Understanding of the complex interactions which occur under anaesthesia, has not reached the level where a program of the fourth class could be devised.

There are two factors which contribute to the difficulty of devising a program of the fourth class for teaching the uptake and distribution of an inhalational anaesthetic. The first is the considerable and unpredictable variation between individuals. The second, and more important factor, is that some of the patient state variables are not observable in man for ethical reasons. Measurements can sometimes be made on animals although it is not always easy to extrapolate the results to man.

When the aim is to teach principles rather than skills, a program of the third class has advantages over one from the fourth class. A realistic model might have several variables changing at the same time, all of which combine to cause the observed output (as with blood pressure control; Guyton, 1971). If these factors are known and can be isolated, a program of the third class can present each effect separately for the student to examine.

Despite the present lack of information in certain critical areas (e.g. CO_2 dynamics under halothane anaesthesia) a program of the third class has been devised for the teaching of the uptake and distribution of the inhalational anaesthetic halothane (it is presented in Chapter 2).

1.5 COMPUTED TOMOGRAPHY

Computed tomography (CT) using X-rays has proved to be a useful tool in diagnostic radiology (cf. Brooks and Di Chiro, 1976). This section briefly introduces the relevant theory.

Figure 1.1 illustrates the coordinate system that is adopted here.

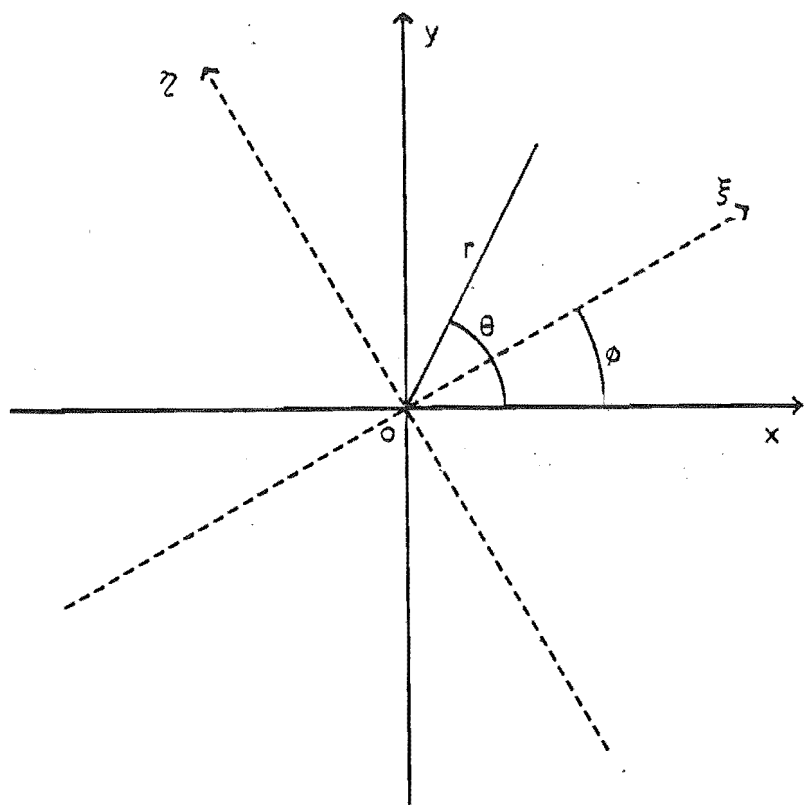


Figure 1.1 Coordinate systems in the Image Plane.

The x,y (cartesian) and r;θ (polar) coordinates are fixed in *image space*. The ξ,η (cartesian) coordinates can be rotated at a variable angle ϕ to the x axis. $\lambda(r;\theta)$ is defined to be the image density in polar coordinates and $\lambda(\xi,\eta)$ is the image density expressed on the rotated cartesian grid. The physical quantity corresponding to density depends on the particular application. For X-ray computed tomography the quantity imaged is the X-ray *linear attenuation coefficient* (Brooks and Di Chiro, 1976).

Photoelectric and Compton scatter interactions contribute to the total linear attenuation coefficient (cf. Alvarez and Macovski, 1976). The photoelectric effect is inversely proportional to the cube of X-ray energy. The energy dependence of Compton scatter can be approximated by the Klein-Nishina function (Alvarez and Macovski, 1976). Both effects depend on the atomic weight, atomic number, and mass density of the material through which the X-rays pass.

The measurements taken from the object are sets of *projections* $f(\xi;\phi)$. The attenuation of the X-ray beam intensity (I) is defined by the formula

$$\frac{\partial I}{\partial \eta} = - \lambda(\xi,\eta)\ I \ . \tag{1.1}$$

If I_o is the intensity of the X-ray beam at the source, the intensity measured at the detector (I_d) is given by

$$I_d = I_o \exp\left(- \int_{-\infty}^{\infty} \lambda(\xi,\eta)\ d\eta\right) \ . \tag{1.2}$$

The infinite limits merely imply that each *ray* of the projection - i.e. each line parallel to the η-axis - traverses the whole of the cross-section. In practice, measurements are made on objects of finite size.

Equation (1.2) can be manipulated to show that it is possible to obtain from measured data the integrated density along lines parallel to the η axis (see Figure 1.1). Thus

$$f(\xi;\phi) = - \ln\left(\frac{I_d}{I_o}\right) = \int_{-\infty}^{\infty} \lambda(\xi,\eta)\ d\eta \ . \tag{1.3}$$

The *projection theorem* (cf. Bates and Peters, 1971) relates the projections to the Fourier transform of the density:

$$\Lambda(\rho;\phi) = \int_{-\infty}^{\infty} f(\xi;\phi)\ \exp(i\ 2\pi\ \rho\ \xi)\ d\xi \tag{1.4}$$

where $i = \sqrt{-1}$. This means that $p;\phi$-space is built up by sets of one-dimensional transforms (Figure 1.2). From this the image density can be obtained by a two-dimensional Fourier transform:

$$\lambda(r;\theta) = \int_{o}^{\infty}\int_{o}^{2\pi} \Lambda(\rho;\phi)\ \exp\left[-i2\pi\rho r\ \cos(\theta - \phi)\right]\ \rho\ d\phi\ d\rho \ . \tag{1.5}$$

There has been some interest in keeping these equations in polar form (cf. Hansen, 1981a,b), in an attempt (apparently unsuccessful) to

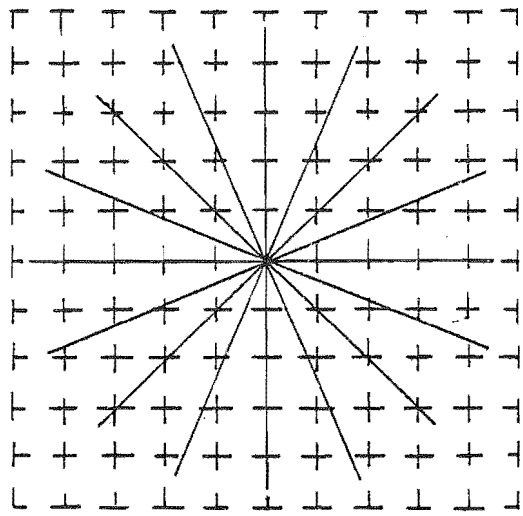


Figure 1.2 $\rho;\phi$-space: the Fourier transforms of the projections (-) lie on radial lines. The transform is required at points on the rectangular grid (--).

develop faster image reconstruction algorithms. However, the usual reconstruction algorithm is found by manipulating these equations as follows. The *modified* (or *filtered*) *projection* is defined:

$$g(\xi;\phi) = \int_{-\infty}^{\infty} |\rho| \, \Lambda(\rho;\phi) \exp(-i2\pi\rho\xi) \, d\rho \, . \qquad (1.6)$$

This is the convolution of the measured projections with a filter whose Fourier transform is $|\rho|$. By substituting this into (1.5), it is found that

$$\lambda(r;\theta) = \int_{0}^{\pi} g(r \cos(\theta - \phi);\phi) \, d\phi \, . \qquad (1.7)$$

The operations specified by (1.6) and (1.7) form the basis of a reconstruction algorithm which can be described as follows. The measured projections are first filtered as in (1.6). Each modified projection is then smeared back across image space at the angle at which it was measured. This is repeated for all measured projections and the results summed. This is the standard reconstruction method in CT and is known as *modified back-projection*. Projections are measured at N discrete angles and each projection has M samples. This means that (1.6) and (1.7), which are in continuous form, have to be replaced by approximate expressions. The consequences of this are discussed in the next section.

1.6 NON-LINEARITIES IN COMPUTED TOMOGRAPHY

It would be desirable to take measurements for X-ray CT in the following way. A mono-energetic pencil-beam of X-rays would scan a stationary object from many different angles. A large number of noise-free measurements would be recorded at each angle. Since none of these conditions applies exactly in practice, reconstructed images tend to exhibit artefacts which can cause serious degradation of image quality.

CT machines have gone through several stages of development since they were first introduced ten years ago. The first machines (conventionally termed 1st generation) effectively irradiate the body with only one ray at a time. The source X-rays are collimated to a pencil-beam which traverses the object and is measured by the detector. There is only one detector. The source and detector translate across the section to measure one projection. The whole configuration then rotates and the procedure is repeated. First generation CT machines have typical scan times of the order of 4-5 minutes (cf. Brooks and Di Chiro, 1976).

In first generation CT scanners, all the rays used to measure one projection are orientated at the same angle. This is known as a *parallel ray* projection (see Figure 1.3).

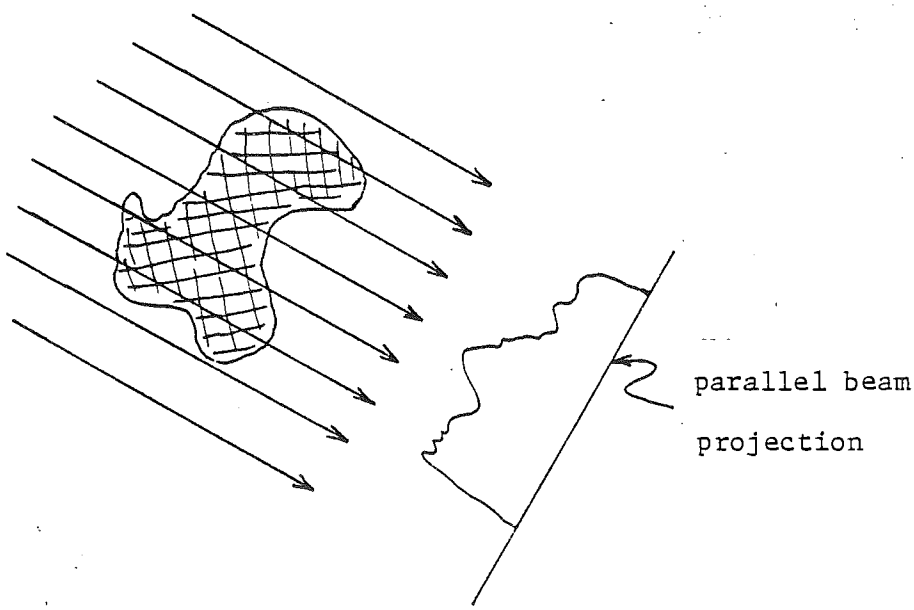


Figure 1.3 Parallel beam projections.

Later model CT machines simultaneously measure several rays. In such a machine the X-rays form a beam which diverges from a point source. This collection scheme produces what is known as a *fan beam* projection (see Figure 1.4).

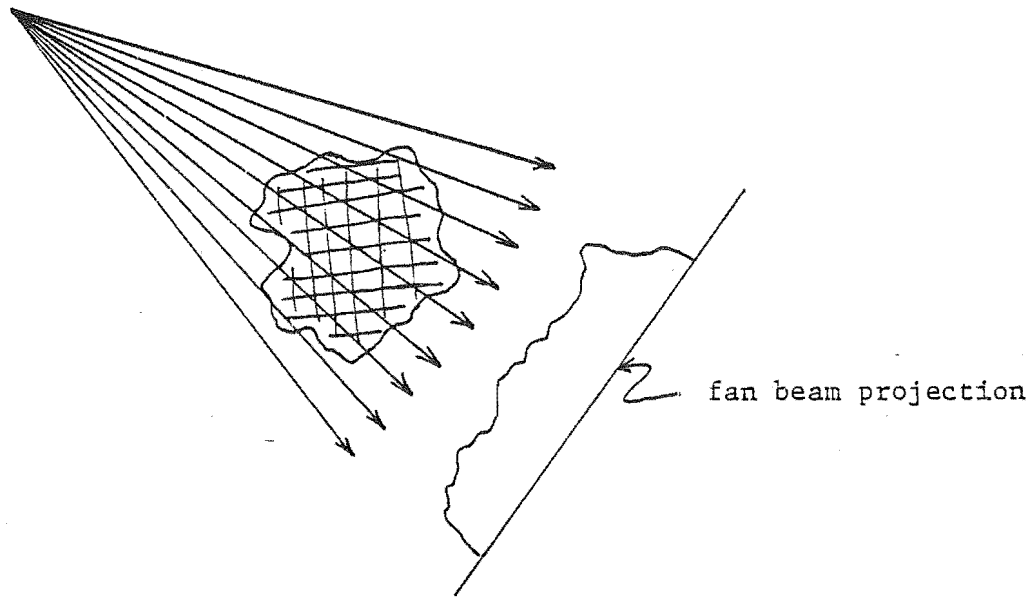


Figure 1.4 Fan beam projection.

Second generation machines are distinguished from the first by the incorporation of multiple (3-60) detectors (cf. Brooks and Di Chiro, 1976). The beam takes the form of a small-angle fan. The source and detectors still follow a translate-rotate scheme, but the scan time is reduced because many rays are measured simultaneously (typical scan times are in the range 20 seconds - 4.5 minutes).

The fan beam covers the whole cross-section in third generation scanners. Thus, the X-ray beam can rotate continuously, since the translational motion is no longer necessary. This gives scan times of the order of 2-20 seconds. The detector array (typically comprised of more than 100 detectors) rotates synchronously with the source.

Fourth generation scanners have a ring of up to 720 detectors spanning 360 degrees. The only motion is the rotation of the X-ray source (cf. Ohio-Nuclear advertisement entitled "Ohio-Nuclear CT Report #1: You can't get Fourth-Generation data out of Third-Generation scanners" which appeared in J. of Computer Assisted Tomography, V1(4), October 1977). These machines have scan times of the order of 2-20 seconds.

In the remainder of this thesis the term 'projection' refers to a parallel ray projection. There is no loss of generality in treating only the parallel ray case, as a complete set of fan beam projections can always be rearranged into parallel beam form (cf. Peters and Lewitt, 1977).

It is usual to express image density on a rectangular grid. Equation (1.5) could form the basis of a reconstruction algorithm once the Fourier transform ($\Lambda(\rho;\phi)$) is known. This would be a computationally efficient operation if a *Fast Fourier Transform* (FFT) could be used (cf. Bergland, 1969). However, use of an FFT requires that the data are known at points on a uniformly spaced grid. Figure 1.2 shows $\rho;\phi$-space for eight equally spaced projections measured with parallel ray geometry. Projections are obtained along radial lines in $\rho;\phi$-space (the values along these radial lines are given by the one-dimensional Fourier transforms of the measured projections - equation (1.4)). Thus, in order to implement a two-dimensional FFT, values must be interpolated on to a rectangular grid (cf. Peters, 1973; Stark *et al.*, 1981a), a cause of large image errors (cf. Peters, 1973).

However, interpolation in Fourier space is avoided when (1.6) and (1.7) form the basis of the reconstruction algorithm. The interpolation required for this method (modified back-projection) is in image space. Reconstruction by this method is found to be less sensitive to interpolation errors.

This section examines the details of practical CT machines to illustrate the causes of artefacts and show how they can be overcome. Two approaches are considered. The first uses a model of the measurement process to characterise the image artefacts. The second approach groups methods which attempt to avoid or remove image artefacts. These methods are based on models of either the measurement process or the image.

1.6.1 ξ Sampling

The image is known to exist within a circle of diameter D. The projection, $f(\xi;\phi)$, is sampled at a spacing $d = D/M$ to give M samples $f(\xi_m;\phi)$ where $1 \leq m \leq M$. The sampled projection $f_s(\xi;\phi)$ is defined by:

$$f_s(\xi;\phi) = d \sum_{m=-\infty}^{\infty} f(md;\phi)\ \delta(\xi - md) \qquad (1.8)$$

where $\delta(\cdot)$ is the Dirac delta function. Only a finite number M of the projection samples are non-zero. The sampled projection has a periodic Fourier Transform since

$$f_s(\xi;\phi) = \int_{-\infty}^{\infty} \left[d \sum_{\ell=-\infty}^{\infty} \Lambda(\rho - \ell/d;\phi)\right] \exp(-i2\pi\rho\xi)\, d\rho\ . \qquad (1.9)$$

Provided $\Lambda(\rho;\phi)$ is negligible for $|\rho| > 1/2d$, the Nyquist frequency, there will be no aliasing error in the projection samples (Lewitt *et al.*, 1978). In this case, (1.6) and the convolution theorem for Fourier Transforms lead to

$$g(jd;\phi) = d \sum_{m=-\infty}^{\infty} f(md;\phi)\, q((j-m)d), \quad j = 0, \pm 1 \ldots \tag{1.10}$$

where the spatial filter $q(\xi)$ is given by

$$q(\xi) = \int_{-\frac{1}{2d}}^{\frac{1}{2d}} |\rho| \exp(-i2\pi\rho\xi)\, d\rho. \tag{1.11}$$

The filter array elements are samples of this function:

$$\begin{aligned} q(kd) &= \frac{1}{4d^2} \quad , \quad k = 0 \\ &= 0 \quad , \quad k \neq 0 \text{ and even} \\ &= \frac{1}{\pi^2 d^2 k^2} \quad , \quad k \text{ odd} . \end{aligned} \tag{1.12}$$

The transform of the filter array consists of the truncated $|\rho|$ characteristic, periodically repeated, as illustrated in Figure 1.5. The signal-to-noise ratio of the measured projections is usually lowest near the folding frequency ($|\rho| = 1/2d$, where d is the sample spacing). Therefore the amplitude of the spectral filter is usually tapered off near odd harmonics of the folding frequency (cf. Lewitt *et al.*, 1978). Also the $|\rho|$ spectral filter has discontinuous derivatives at the folding frequency harmonics leading to overshoot in the reconstruction of sharp transitions in the object (e.g. Gibbs' phenomenon). Figure 1.5 also illustrates the filter of Shepp and Logan (1974) which improves these aspects, at the expense of a small loss of resolution compared to the periodic $|\rho|$ filter.

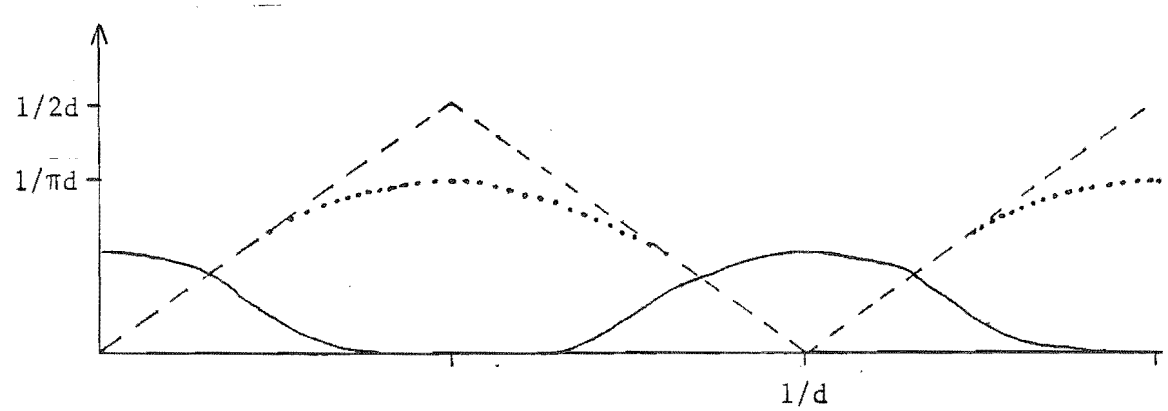


Figure 1.5 The periodic spectra of the sampled projection (——), the $|\rho|$ filter (---), and the Shepp-Logan (1974) filter (•••).

The first step in the modified back-projection method is to filter the projections. The convolution for sampled projections is given in (1.10). This can be implemented straightforwardly as a convolution sum. An alternative procedure is to first calculate the Fourier transform of the projection. This is then multiplied by the Fourier transform of the filter array. The inverse Fourier transform gives the modified projection. All of this can be conveniently effected with the aid of the FFT algorithm.

1.6.2 ϕ Sampling

Projections are measured at N angles, which are usually equally spaced throughout an interval of π radians. The back-projection operation (1.7) is expressed as a continuous integral in ϕ. However there are only a limited number (N) of samples spanning the range of the integral. Equation (1.7) therefore has to be evaluated by some numerical procedure. For example, (1.7) can be re-written as

$$\lambda(r;\theta) = \sum_{n=1}^{N} \int_{\frac{\pi}{N}(n-\frac{1}{2})}^{\frac{\pi}{N}(n+\frac{1}{2})} g\big(r\cos(\theta-\phi);\phi\big)\, d\phi \ . \qquad (1.13)$$

However, there is only one projection in the range $\phi = (n-\frac{1}{2})\pi/N$ to $\phi = (n+\frac{1}{2})\pi/N$. Brooks *et al.* (1978) suggest converting the ϕ integral into an integral involving a radial coordinate:

$$\lambda(r;\theta) \simeq \sum_{n=1}^{N} \frac{\Delta\phi}{S_{n+1}-S_n} \int_{S_n}^{S_{n+1}} g\big(r\cos(\theta-\phi_n);\phi_n\big)\, ds \qquad (1.14)$$

where $S_n = r\cos\big((n-\frac{1}{2})\pi/N-\theta\big)$, $\Delta\phi = \pi/N$ and $\phi_n = n\pi/N$. The usual approach, however, is to approximate each of the integrals in (1.13) by the zeroth-order Newton-Cotes formula. This means that (1.13) becomes a sum containing one value from each projection:

$$\lambda(r;\theta) = \sum_{n=1}^{N} g\big(r\cos(\theta-\phi_n);\phi_n\big) \ . \qquad (1.15)$$

It is found that (1.14) and (1.15) behave similarly. However for small N, (1.14) gives marginally superior results (Brooks *et al.*, 1978).

There is aliasing error in the reconstruction unless there is sufficient ϕ sampling. The finite number of projections also causes an interpolation error when the back-projection is evaluated by (1.15). This error is analysed by Smith, Peters and Bates (1973) for the case when the projections are continuous functions of ξ. They calculate an

expression for the error, which they call 'clutter', for the reconstruction of a gaussian density distribution. Lewitt, Bates and Peters (1978) have extended this result to include the effect of ξ sampling. Kowalski and Wagner (1977) use an experimental approach to link finite ξ and ϕ sampling. Similar results are obtained by Huesman (1977).

One way to minimise the X-ray exposure given to a patient is to take few projections. This means that an image has to be formed from only a small set of projections and so reconstruction techniques which produce good images for small N can help to reduce the X-ray dose.

Such techniques are also necessary when imaging the beating heart. This is because the two main approaches used to image the beating heart both produce only small sets of projections to reconstruct each phase of the heart cycle. The Dynamic Spatial Reconstructor (DSR) built at the Mayo Clinic uses 28 X-ray tubes which irradiate the patient sequentially (Robb *et al.*, 1979). Projections are measured by 28 TV cameras which view a fluorescent screen which detects the X-rays. This machine can simultaneously measure 60 cross-sections in 1/60 of a second. However there are only 28 projections available to reconstruct each cross-section at each instant.

Another way to image the beating heart is to synchronise the projections with the patient's E.C.G. (electrocardiogram) while the breath is held. If the scan is taken over about 10 seconds there will be of the order of 10 heart beats. All the projections which occur at the same phase of the E.C.G. are grouped together. These can then be used to reconstruct an image of the heart corresponding to that phase of the E.C.G. (McKinnon and Bates, 1981). Again there is the problem of reconstructing an image from few projections.

Sato *et al.* (1981) have recently suggested that the use of *a priori* image information can improve the reconstruction when there are few measured projections. This is investigated further in Chapter 4.

1.6.3 Missing sector of projections

Projections are required to be uniformly spaced throughout π radians so that the back-projection operation (1.15) can be applied. However, there are some situations when it is desirable, or necessary, to have unevenly-spaced projections. In particular, there are two

situations in which a range (*sector*) of projections is not measured. The first can occur if the cross-section of interest contains some material of high density. For example, the section may contain an artificial joint. Thus projections cannot be measured (or are very noisy) in certain directions. The second reason for the interest in reconstruction from a limited range of projection angles is that it offers the possibility of reducing the X-ray dose to the patient. The smaller the number of projections that are measured, the less X-ray dose the patient is subjected to.

When there is a missing sector of projections, (1.15) cannot be applied straightforwardly. The usual approach is to attempt to estimate the missing projections by extrapolating into the missing sector of projection space. This enables the conventional image reconstruction algorithms to be applied. Inouye (1979) extrapolates into the missing region by fitting a Fourier series (cf. §4.1.2) to the measured projections. However, this scheme has proved sensitive to noise in the measured projections. Similar results have been obtained by Wagner (1979b) who uses a cubic spline to estimate the projections in the missing sector. Tam and Perez-Mendez (1981) and Sato *et al.* (1981) use a scheme based on the well-known Gerchberg (1974) algorithm. The missing projections are estimated with the aid of *a priori* knowledge of the image extent. The details of this algorithm are presented in §4.3.

1.6.4 Incomplete projections

A projection which is measured over an insufficient range of ξ to cover the entire object is known as *incomplete*. As with a missing sector of projection angles, incomplete projections could be measured when the object contains some material of high density. For example if the section contains an artificial joint, each measured projection has a missing region (cf. Hinderling *et al.*, 1979). Incomplete projections are described as either hollow or truncated depending on whether the interior or edges of the projections are missing (Lewitt and Bates, 1978a). The artefacts introduced by straight reconstruction from these incomplete projections are well known (Lewitt and Bates, 1978a; Gore and Leeman, 1980). Exact reconstruction is theoretically possible for hollow projections (Lewitt and Bates, 1978a), although the computational approach for both truncated and hollow projections is to first complete the projection by some approximate but plausible scheme (Lewitt and Bates, 1978b and c).

The image can then be reconstructed by a conventional scheme such as modified back-projection (1.10 and 1.15). The outline of the body must be known if the truncated projections are to be completed successfully. This outline (commonly called the *extent*) can be obtained by physical measurement (cf. Wagner, 1979a). Another approach relies on the assumption that the parts of the body which do not contribute to a particular incomplete projection are only composed of water (Heuscher and Carlson, 1979). In this method a curve is fitted to the end of the truncated projection so that the extended part is equivalent to the projection of a circular region consisting of water only.

Others (Nalcioglu *et al.*, 1979; Stanton and Tretiak, 1981) suggest an alternative to measuring incomplete projections. In their method, the whole projection is measured, but only the region of interest receives the full X-ray dose. This circumvents the difficulty of extending incompletely measured projections while reducing the X-ray exposure for the patient.

1.6.5 Detector sensitivity

Variation of detector sensitivity during a CT scan causes image artefacts. One way to reduce the effect of drifting detector gain and offset is to recalibrate the detectors during the scan. The artefact due to sensitivity variation depends on the scanner geometry. Figure 1.6 shows the ray paths viewed by one particular detector of a third generation scanner.

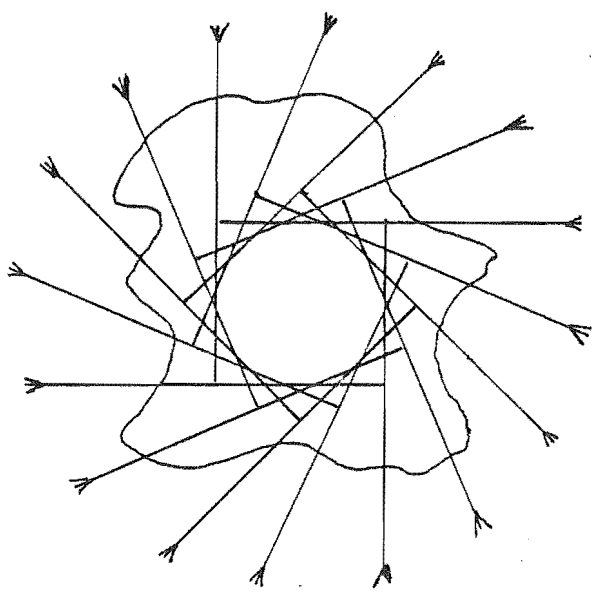


Figure 1.6 The ray paths measured by one particular detector in a third generation scanner.

In this case all these ray paths go through the object. However with a fourth generation scanner, each detector first sees air, then the object, then air again as the X-ray tube rotates. This means that each detector can be calibrated during the scan. However, the patient has to be removed before calibration of the detectors in a third generation scanner. Figure 1.6 shows that a particular detector in a third generation machine is always in the same relation to the X-ray tube source. Thus detector drift causes circular artefacts in third generation machines (cf. Kowalski, 1977a). However, each detector in a fourth generation scanner receives rays that successively pass through all regions of the object. Thus any artefact due to detector drift covers the entire image. Consequently, detector stability is less critical in fourth generation scanners. One way round this difficulty with third generation scanners, suggested by Kowalski (1977b), is to modify them so that the detectors move relative to the source during scanning.

1.6.6 Beam shape

Figure 1.7 illustrates the usual shape of the X-ray beam used in Computed Tomography (cf. Brooks and Di Chiro, 1976).

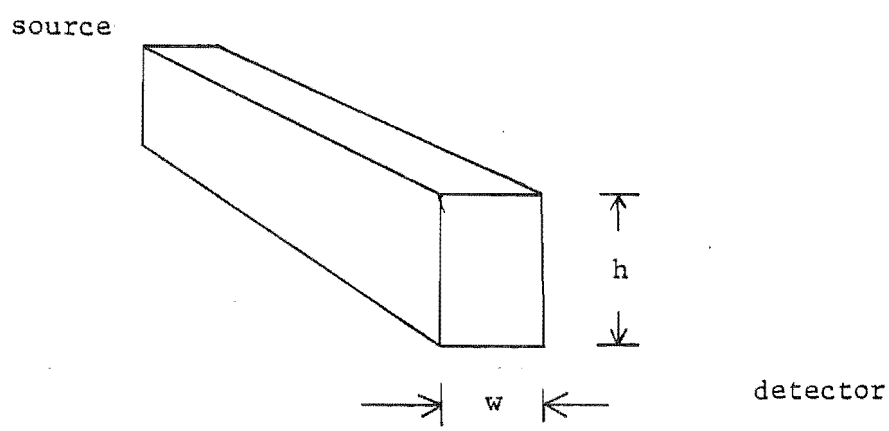


Figure 1.7 Beam shape as determined by the source and detector collimators. The slice thickness is 'h' and the beam width is 'w'.

The beam shape is determined by the source and detector collimators. The larger dimension (h) is at right-angles to the plane of the cross-section. This dimension is referred to as the *slice thickness*.

The *beam width* w is the dimension of the beam in the plane of the cross-section. The slice thickness is dictated by the need to capture sufficient photons to keep statistical fluctuations at a low level. However the finite thickness means that the image is effectively blurred at right-angles to the image plane. Therefore, resolution perpendicular to the image plane is much worse than that in the image plane. Glover and Eisner (1979) and (1980) analyse the resolution limits set by finite detector size and spacing and source distribution. Cormack (1978) shows that the consequence of sampling with beams of finite width is to under-sample the available information. However, it is possible to compensate for collimator width. This can be done by deconvolving the collimator response from the measured projections (cf. Bracewell, 1977). Peters (1974) suggests using a Wiener filter to remove this detector sensitivity profile.

1.6.7 Beam energy

The aim of CT is to label every point within a cross-section by a number which is characteristic of the tissue occupying that point. When the measurements are performed with monoenergetic photons (e.g. from a radionuclide), all points within the body are sampled at the same energy, which is equal to the energy of the incident photons. A CT reconstruction using monoenergetic photons yields an image with the attenuation coefficients of all materials corresponding to the same energy. However, CT machines usually employ polyenergetic X-rays in order to obtain high flux rates. Since the attenuation of photons at lower energies is greater than at higher energies, the effective energy of the beam shifts upwards as the beam travels through the medium (cf. Nalcioglu and Lou, 1979). This phenomenon is known as *beam hardening*. Beams in different directions pass through different tissue and have different spectra when detected. This makes it difficult to assign a single value for the attenuation coefficient of a point in the body. In soft tissue, beam hardening causes distortion of tissue density. This artefact (called *cupping*) is evident as raised image density towards the edges of the image (cf. Joseph and Spital, 1978). Beam hardening also causes streaks from dense objects such as bone (Duerinckx and Macovski, 1978). A comprehensive review of beam hardening compensation methods is presented by Stonestrom *et al.* (1981). All the correction schemes use *a priori* assumptions about the image. These assumptions are embodied in a model for the composition of biological

tissues. For example the tissue can be modelled as a single material such as water. Such models are inferior to those which assume that the tissue is composed of a (variable) combination of two materials. However, in a practical implementation for beam-hardening correction, it may be sufficient to use a look-up table or a low order polynomial to convert the polyenergetic projections to their equivalent monoenergetic projections (cf. Herman, 1979).

1.6.8 Noise

When an X-ray photon passes through a material it may interact with it by scattering off an electron. (An inelastic collision between an X-ray photon and an electron is called *Compton scatter*.) The photon is then sent off at a new angle and with reduced energy (cf. Stonestrom and Macovski, 1976). Harding (1981) has recently studied the use of this Compton scatter to image the material. However, it is undesirable to measure this scatter in CT. The detectors usually have a collimator to exclude X-rays which have not come directly through the material. Any scattered X-rays which do reach the detectors are effectively random noise on the measured intensity. The first step in the processing of the measurements (1.3) is to form the projections as minus the logarithm of the measured intensity. This means that anywhere that I_d is small, the projection is large. Thus random noise on I_d gives large projection noise where the projection is large (I_d small), and small projection noise where the projection is small (I_d close to I_o). Hence detector noise causes object-dependent artefacts. The effect of noise in CT is analogous to film grain noise in optical imaging systems. There, the non-linear detector sensitivity causes signal dependent noise (cf. Andrews and Hunt, 1977, Chapter 2). Stonestrom and Macovski (1976) have used computational models to show that the artefacts due to noise are negligible in CT if the scatter level is less than 1% of the detector signal. Similar results are obtained by Tofts and Gore (1980) who simulate the effect of adding a uniform scatter to the measurements. Dark current of the detectors and quantum noise also cause object-dependent errors (Kowalski, 1977c). Duerinckx and Macovski (1979) point out the similarity between the artefacts due to noise and polychromacity. They call these artefacts *non-linear shadows*. In these situations the projections can be *inconsistent* if the object is not circularly symmetric (cf. Smith, Peters and Bates, 1973 for a definition of consistency).

1.6.9 Patient motion

One of the basic postulates of CT is that all scans are geometrical projections of the same object. However this is violated if there is any movement of the object. Figure 1.8 summarises the types of physiological motion which can cause blurring if they occur while the projects are being measured (cf. Alfidi *et al.*, 1976). Current CT machines can take a body scan in as little as 2 seconds. A patient can voluntarily suspend respiration comfortably for up to 12 seconds. Thus respiratory motion is not necessarily a difficulty with chest scans. However, involuntary muscular motion and cardiovascular motion do cause blurring in routine CT. Kowalski and Wagner (1977) examine the particular case where the whole cross-section moves in the reconstruction plane. In this case the projections can be shifted to remove the artefact. Motion artefacts are simulated by Alfidi *et al.* (1976) using a physical model of the types of motion found in man. Their conclusion is that, generally, motion causes loss of contrast and spatial resolution.

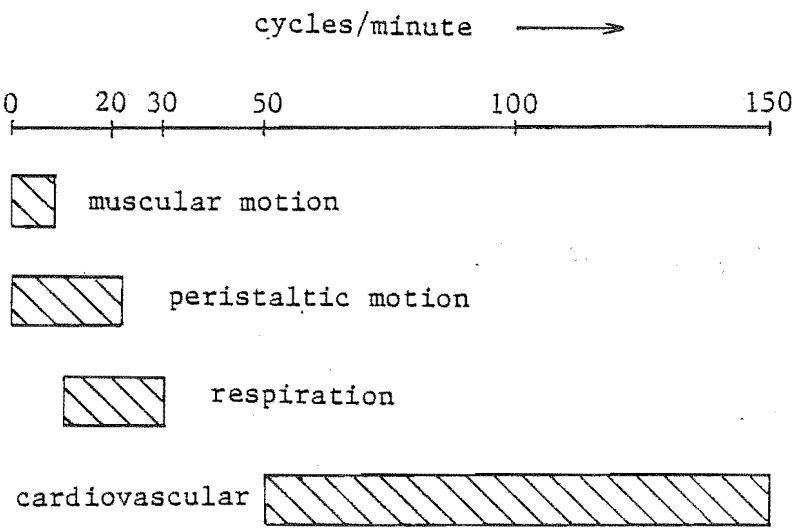


Figure 1.8 Approximate frequencies of periodic and aperiodic physiological motion.

Study of these artefacts is extended in Chapter 3. There, models are developed for different types of movement. These are used to quantify the loss of resolution caused by motion artefacts.

2. TEACHING HALOTHANE UPTAKE AND DISTRIBUTION

2.1 INTRODUCTION

A computer aided learning system has been under development at the Christchurch Clinical School over the last three years. The system administers multi-choice questions and patient management problems. It also contains simulation programs from the 'Mac' series developed at McMaster University in Hamilton, Ontario, Canada (Dickinson, 1977).

Experience at the Christchurch Clinical School with the 'Mac' series has shown the value of simulations of physiology and pharmacology for teaching medical students. These simulations have proved to be useful teaching aids in a tutorial environment. This has led to the development of a simulation program for the uptake and distribution of an inhalational anaesthetic. The program specifically models halothane, although the principles on which the model is based also apply to other drugs. The primary intention is not to teach the skills required for anaesthetic administration, but rather to help students understand the pharmacokinetics involved. This program falls into the third class of CAL programs discussed in §1.4. Two papers covering the work reported in this chapter are to appear in "Anaesthesia" in 1982.

The principles of halothane anaesthesia are introduced in §2.2. In §2.3 the model for the uptake and distribution of halothane is presented. The interaction between computer and student is illustrated in §2.4. Some illustrative simulations from a teaching session are examined in §2.5, and §2.6 presents the results of an evaluation which aimed to determine the usefulness of this program as a teaching tool. The practical value of the model and the CAL program are discussed in §2.7.

2.2 HALOTHANE ANAESTHESIA

Halothane (C_2 HCl Br F_3) is a volatile anaesthetic agent. It is used for general anaesthesia since it produces loss of consciousness, a reduction of muscle tone, and amnesia. Halothane is administered by inhalation and is often used in conjunction with other anaesthetic gases,

e.g. Nitrous Oxide. Other drugs such as muscle relaxants and analgesics are frequently given in association with halothane. Throughout this chapter, it is assumed that halothane is the only gas taken up by the patient. All other gases are considered as inert and are excluded from the model. Considering halothane as the only pharmacologically active agent enables the student to concentrate on its effects.

Halothane is administered to the patient as a vapour through a set of breathing tubes which are known as the *delivery system* or *circuit*. Halothane is introduced into this circuit from an accurately calibrated vaporiser which contains a reservoir of halothane stored as a liquid. This vaporiser and its associated delivery apparatus have adjustments which alter the flow of gas into the circuit and the fraction of halothane which is vaporised into the flow. The fractional concentration of halothane is controlled by a dial on the vaporiser to an accuracy of 0.1%.

There are several different types of delivery systems used to connect the patient to the anaesthetic machine. The most commonly used ones are termed 'rebreathing' or 'circle' since exhaled gases can recirculate to the patient. This means that the anaesthetic agent is conserved and that the potential for atmospheric pollution is reduced. Rebreathing systems contain an absorber to remove CO_2 from the expired gases and a flexible reservoir (usually a rubber bag) which allows visual assessment of ventilation. The particular anaesthetic delivery system modelled in this chapter is known as 'circle H' (Eger, 1974, Chapter 13). This is a rebreathing circuit which is designed to allow preferential rebreathing of dead-space gas, i.e. that gas breathed by the patient which does not get to the site of gas exchange - the alveolus. This brief introduction to anaesthetic circuits does not cover details, such as the placement of valves and fresh gas inflow to obtain the stated functions. The main advantage of a 'circle H' system is that dead-space gas is rebreathed first. Thus alveolar gas is spilled to the atmosphere in preference to dead-space gas which generally has a higher concentration of halothane because there has been no opportunity for halothane absorption. This aids circuit economy.

Before describing the model developed for the CAL program it is instructive to consider the anaesthetist's role. This is to control the rate of flow of halothane to obtain and maintain an appropriate brain concentration of halothane. The anaesthetist adjusts the anaesthetic machine in accordance with clinical judgement based on the actions of

the surgeon, feedback of the state of the patient, and the anaesthetist's experience. The patient variables monitored by the anaesthetist are such things as heart rate, blood pressure and colour of the patient.

Figure 2.1 shows the structure of the model for halothane uptake and distribution, which is discussed in detail in §2.3.

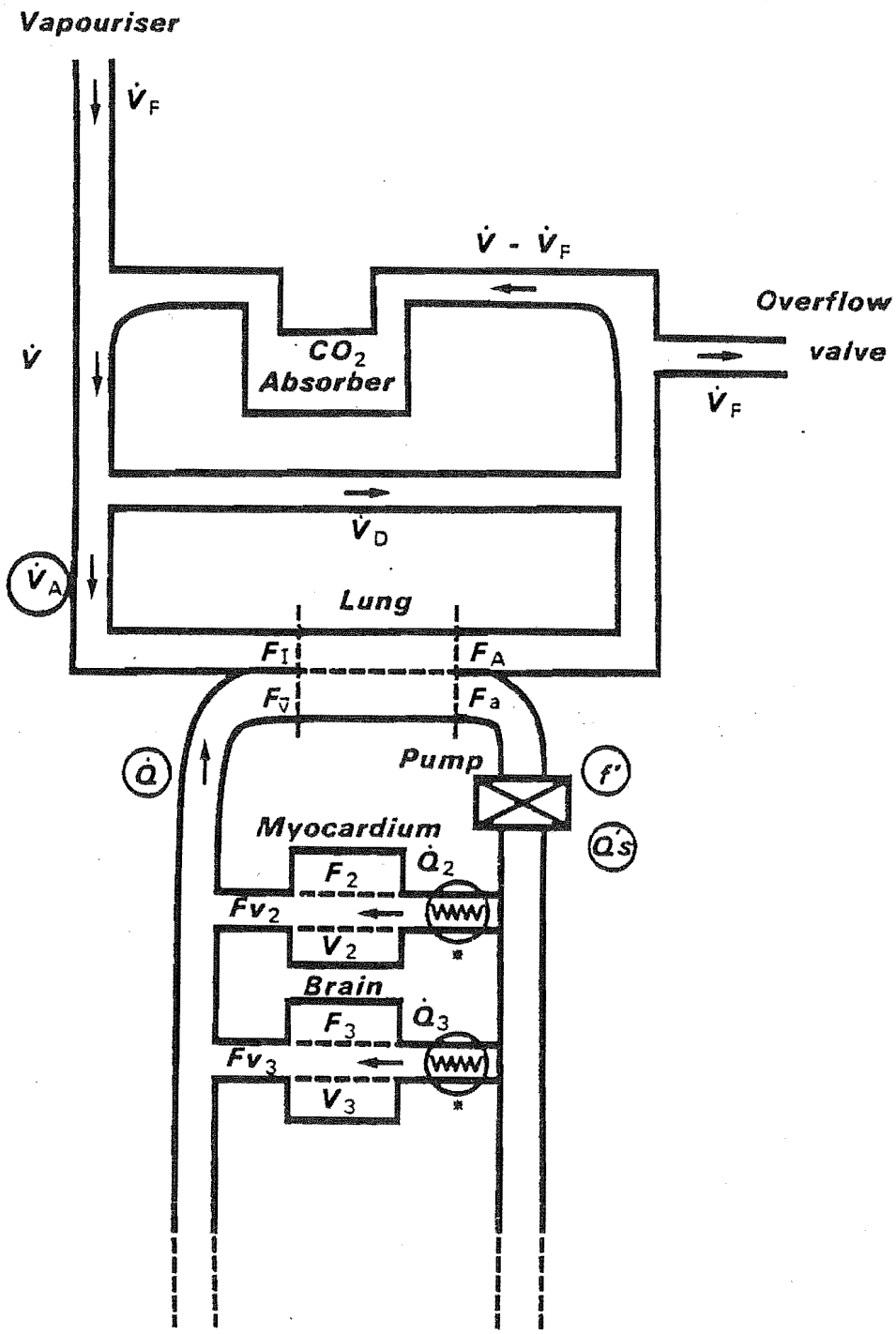


Figure 2.1 Model for the anaesthetic circuit and tissues. The symbols and their meaning are listed in Table 2.1. Those variables which are circled are controlled from within the model. The symbols with the '*' beside them are the variable blood-flow resistances. The diagram does not include tissues 4-8, to conserve space.

The top of this diagram illustrates the anaesthetic delivery system part of the model which is in the gas phase. Halothane is one of the gases in this mixture. The lower part of the diagram (from the air-blood interface in the lungs, down) illustrates the system of distribution of the anaesthetic round the body. Halothane is dissolved in the blood and tissues.

2.3 THE MODEL

Following the conventions of respiratory physiology (Pappenheimer *et al.*, 1950), all respiratory volumes (V) and flows ($\dot{V}$) are expressed in this chapter at BTPS (Body Temperature and Pressure Saturated). Therefore, for consistency, the volumes and flows in the rebreathing system are also expressed at BTPS. The student specifies fresh gas flow as it would be read from the rotameter at ATPD (Ambient Temperature and Pressure Dry) but the program adds the flow of halothane vapour derived from the vaporiser and converts from ATPD to BTPS. Concentrations of halothane (F) are expressed as fractions of total gas rather than the conventional fractions of the dry part. The symbols used in the description that follows are listed in Table 2.1.

Anaesthetic circuit

The simulation models a rebreathing type anaesthetic circuit with carbon dioxide absorption, described by Eger (1974) as 'circle H' and depicted diagrammatically in Figure 2.1. It is assumed that there are no losses from the circuit by leaks or by diffusion of halothane into the rubber tubing. There are three separate situations to be considered.

(i) There is no rebreathing if $\dot{V}_V > \dot{V}$. In this case

$$F_I = F_F \tag{2.1}$$

and

$$\frac{d\,F_b}{dt} = 0 \,. \tag{2.2}$$

(ii) Figure 2.1 shows the case where there is some rebreathing ($\dot{V}_F < \dot{V}$). The flow rate of halothane vapour entering the breathing system = $F_F\,\dot{V}_F$, and the flow rate of halothane from the absorber = $F_b(\dot{V} - \dot{V}_F)$. Therefore

$$F_I = F_F \frac{\dot{V}_F}{\dot{V}} + F_b \frac{\dot{V} - \dot{V}_F}{\dot{V}} \tag{2.3}$$

Table 2.1 Symbols

F_A	-	volume fraction of halothane in alveolar gas at BTPS
F_a	-	volume fraction of halothane in the gas with which the arterial blood is in equilibrium at BTPS
F_b	-	volume fraction of halothane in the gas in the CO_2 absorber at BTPS
F_F	-	fraction of halothane in the fresh gas flow at BTPS (converted from the dry gas value dialled by the student)
F_I	-	inspired volume fraction of halothane at BTPS
F_i	-	volume fraction of halothane at BTPS which would occur in a small volume of gas brought into equilibrium with tissue compartment 'i'
F_v	-	volume fraction of halothane at BTPS which would occur in a small volume of gas brought into equilibrium with mixed venous blood
F_{vi}	-	volume fraction of halothane at BTPS which would occur in a small volume of gas brought into equilibrium with blood leaving tissue 'i'
f'	-	normalised heart rate
G_i	-	conductance of tissue 'i' - litres min^{-1} $mmHg^{-1}$
ΣG	-	sum of the compartmental conductances - litres min^{-1} $mmHg^{-1}$
K_i	-	blood flow redistribution factor - litres min^{-1} $mmHg^{-1}$
P_a	-	mean arterial blood pressure - mmHg
P_d	-	diastolic pressure - mmHg
P_i	-	partial pressure of halothane in tissue 'i' - mmHg BTPS (= 760 F_i)
P_p	-	pulse pressure - mmHg
P_s	-	systolic pressure - mmHg
$\dot{Q}$	-	cardiac output - litres min^{-1}
$\dot{Q}_i$	-	blood flow through tissue 'i' - litres min^{-1}
Q_s'	-	normalised stroke volume
$\dot{V}$	-	total ventilation at BTPS
$\dot{V}_A$	-	alveolar ventilation at BTPS - litres min^{-1}
$\dot{V}_A'$	-	normalised alveolar ventilation
$\dot{V}_D$	-	dead space ventilation at BTPS - litres min^{-1}
$\dot{V}_F$	-	fresh gas flow at BTPS (converted from the value set by the student at ATPD) - litres min^{-1}
V_b	-	absorber volume - litres
V_i	-	volume of tissue 'i' (see Table 2.2) - litres
V_L	-	volume of lung at BTPS - litres
λ_b	-	blood-gas partition coefficient
λ_i	-	tissue-blood partition coefficient for tissue 'i'

NOTE: Most of the above variables are time varying, but for convenience they have not been given an explicit time subscript in the model description.

Anaesthetic uptake by the walls of the breathing system is neglected. Then, if $\dot{V}_A < \dot{V}_F < \dot{V}$, and because dead space gas is preferentially rebreathed in the circuit modelled here (described in §2.2), halothane enters the absorber at $F_I(\dot{V} - \dot{V}_F)$ and leaves at $F_b(\dot{V} - \dot{V}_F)$. Therefore

$$\frac{d\,F_b}{dt} = \frac{(\dot{V} - \dot{V}_F)\,(F_I - F_b)}{V_b} \quad . \tag{2.4}$$

(iii) If $\dot{V}_F < \dot{V}_A$, all the dead space gas and some of the alveolar gas is rebreathed. Halothane enters the absorber at $\dot{V}_D F_I - (\dot{V}_A - \dot{V}_F)\,F_A$. Therefore

$$\frac{d\,F_b}{dt} = \frac{\dot{V}_D(F_I - F_b) + (\dot{V}_A - \dot{V}_F)\,(F_A - F_b)}{V_b} \quad . \tag{2.5}$$

The inspired halothane level is still given by equation (2.3).

Lung model

There is assumed to be equilibrium between alveolar and arterial pressures of anaesthetic vapour, i.e. $F_A = F_a$. The lung tissue volume of 0.5 litres and arterial blood volume of 1.03 litres (equivalent to 1.2 litres and 2.5 litres respectively of gas at BTPS) (Mapleson, 1973), are effectively part of the lung compartment for this kind of model and have been added to a lung air volume of 2.75 litres (2.5 litres FRC (Mapleson, 1973) plus half a tidal volume of 500 ml (Cowles *et al.*, 1971)) to give an equivalent lung volume (V_L) of 6.45 litres. The partition coefficient for lung tissue is assumed to be the same as the blood-gas partition coefficient (see Table 2.2). Equating gas flows into the lung

$$V_L \frac{d\,F_A}{dt} = (F_{\bar{v}}\,\lambda_b\,\dot{Q} + F_I\,\dot{V}_A) - (F_A\,\dot{V}_A + F_a\,\lambda_b\,\dot{Q}) \quad . \tag{2.6}$$

Tissues and blood flow

Halothane is taken up from the anaesthetic circuit and is distributed to seven compartments - heart, brain, liver, kidney, muscle, poorly perfused tissue (ppt.) and fat. The values given by Mapleson (1973) for tissue mass have been followed except that 'kidney' also includes adrenals and thyroid, 'poorly perfused tissue' consists of red marrow and non-fatty subcutaneous tissue, and 'liver' includes prostate, salivary glands, eyes, thymus, testes, and spinal cord. As suggested by Mapleson (1973), 16% of cardiac output passes through a peripheral shunt.

Table 2.2 Values used in the program

Compartment	Volume (litres)	Partition coefficient	Fractional blood flow*	Redistribution factor †
1: lung	6.45	$\lambda_b = 2.4$	---	---
2: heart	.28	$\lambda_2 = 2.9$	3.7%	$K_2 = 0$
3: brain	1.43	$\lambda_3 = 2.7$	12.3%	$K_3 = .64$
4: liver	3.91	$\lambda_4 = 2.5$	24.5%	$K_4 = 0$
5: kidney	.32	$\lambda_5 = 1.5$	21.4%	$K_5 = -.42$
6: muscle	30.25	$\lambda_6 = 2.5$	10.5%	$K_6 = -.14$
7: fat	12.84	$\lambda_7 = 65$	4.7%	$K_7 = 0$
8: ppt.	7.14	$\lambda_8 = 2.3$	6.9%	$K_8 = 0$
9: absorber	10	---	---	---
10: shunt	---	---	16%	$K_{10} = 0$

The sources for these data are Mapleson (1973) for the tissue volumes and blood flows, Steward *et al.* (1973) for the partition coefficients, and Smith *et al.* (1972) for the redistribution data and the partition coefficient for ppt. Lowe (1972) (Table 12) gives the absorber, together with the associated breathing tubing, a volume of 10 litres.

* These are initial fractions of cardiac output, they will change with redistribution.

† The units for the redistribution factors are litres min^{-1} $mmHg^{-1}$.

No intrapulmonary or extrapulonary right-to-left shunt is included and there is assumed to be no ventilation/perfusion mismatch in the lungs. Thus gas tensions in arterial blood, alveolar gas and lung tissue are all in equilibrium. A standard 70 kg man (Mapleson, 1973) is understood here, although "he" is not intended to represent a real patient, since the aim of this model is to teach principles rather than skills. Quantification for the model is given in Table 2.2.

The tissue level of halothane (F_i) is related to the arterial level (F_a) and to that in the venous blood draining the tissue (F_{vi}) by the Fick equation (cf. Eger, 1974, p. 85):

$$V_i \lambda_i \frac{d F_i}{dt} = F_a \dot{Q}_i - F_{vi} \dot{Q}_i . \qquad (2.7)$$

It is assumed that the partial pressure of halothane in each tissue is the same as that in the venous blood draining it. The mixed venous fraction of halothane is given by the sum of the flow-weighted levels in venous blood leaving the tissues:

$$F_{\bar{v}} = \frac{\dot{Q}_2 F_{v2} + \ldots + \dot{Q}_8 F_{v8} + \dot{Q}_{10} F_a}{\dot{Q}} . \qquad (2.8)$$

The seven tissue compartments are numbered 2-8 (Table 2.2). The blood flow through the peripheral shunt, $\dot{Q}_{10}$, has a halothane concentration F_a.

Ventilation

Continuous (non-cyclical) ventilation is used since the program aims to demonstrate changes which occur on a time scale of minutes, rather than those due to breath-to-breath variations. Thus no attempt is made to simulate the effect of halothane on respiratory rate or tidal volume. The respiratory effects are expressed in terms of alveolar ventilation only. Dead space ventilation is assumed constant at the level specified by the student when setting up the simulation. The influence of this dead space is shown in Figure 2.1. It is shown by Munson *et al.* (1973) that alveolar ventilation is depressed by the halothane concentration in the brain. Following Munson's argument, normalised ventilation is given by

$$\dot{V}_A' = 1 - \frac{900 \frac{P_3}{760} + \left(100 \frac{P_3}{760}\right)^3}{900 \frac{19}{760} + \left(100 \frac{19}{760}\right)^3} \qquad (2.9)$$

where P_3 is the brain partial pressure in mmHg. Actual ventilation is calculated as the product of V_A' and the value for V_A given by the student.

Stroke volume

The way in which Munson *et al.* (1973) modelled cardiac output is here applied only to stroke volume to allow for the long-term recovery of cardiac output by increased heart rate as suggested by Deutsch *et al.* (1962). Stroke volume is depressed by the partial pressure of halothane in the heart muscle (P_2).

$$Q_s' = 1 - \frac{900\,\frac{P_2}{760} + \left(100\,\frac{P_2}{760}\right)^3}{900\,\frac{22.8}{760} + \left(100\,\frac{22.8}{760}\right)^3} \qquad (2.10)$$

Heart rate

There is an observed initial drop in heart rate with halothane anaesthesia (Eger *et al.*, 1971), which is followed by a slow rise (Eger *et al.*, 1971, and Deutsch *et al.*, 1962).

The drop in heart rate is simulated by making the fall proportional to arterial halothane concentration (giving a rapid change) except that for concentrations greater than 1 MAC, the depression is held at about 12% (MAC is defined by Eger (1974) to be the minimum alveolar concentration of anaesthetic at 1 atmosphere that produces immobility in 50 percent of patients exposed to a noxious stimulus).

The slow rise of heart rate is modelled empirically by making heart rate linearly dependent on the halothane level in tissue #6 (muscle). Muscle is chosen because its time constant most closely matches the rate of rise of the average results obtained by Deutsch *et al.* (1962). Thus

$$f' = 1 + 40\,F_6 \,. \qquad (2.11)$$

No conclusion about the mechanisms involved is implied by this - it is only valid in as far as the model follows Deutsch's data. Cardiac output is calculated as the product of f', Q_s', and the value typed in by the student for cardiac output.

Blood flow distribution

Total blood flow (cardiac output) is calculated as the product of heart rate and stroke volume. The fraction of this blood flowing to each tissue is determined by the ratio of that tissue's conductance to the sum of all the tissue conductances. (A tissue's conductance is the reciprocal of its blood flow resistance.) The changes to the distribution of blood flow during halothane anaesthesia (Smith *et al.*, 1972) are modelled by considering the compartmental conductances to be linearly related to the arterial halothane concentration:

$$G_i(F_a) = G_i(0) + K_i\,F_a \,. \qquad (2.12)$$

The pressure-flow relationship is

$$P_{\bar{a}}(F_a) = \frac{\dot{Q}(F_a)}{\Sigma G(F_a)} \qquad (2.13)$$

where the total conductance

$$\Sigma G(F_a) = G_2(F_a) + \ldots + G_8(F_a) + G_{10}(F_a) \; . \qquad (2.14)$$

The flow through tissue 'i' is

$$\dot{Q}_i(F_a) = G_i(F_a)\, P_{\bar{a}}(F_a) \; . \qquad (2.15)$$

Once $G_i(0)$ and K_i are known, (2.12) can be used to calculate $G_i(F_a)$. Equations (2.13) and (2.14) can then be used to calculate arterial pressure once cardiac output has been obtained from (2.10) and (2.11).

Values for K_i and $G_i(0)$ were determined for each tissue from the data given by Smith *et al.* (1972) for equilibrium at 0% and 2% end tidal halothane. For example in the brain (compartment #3) $\dot{Q}_3(0)$ = .75 ℓ/min with normal arterial pressure $P_{\bar{a}}(0)$ = 93 mmHg. Therefore, from (2.15)

$$G_3(0) = \frac{.75}{93} = .0081 \text{ litres min}^{-1} \text{ mmHg}^{-1}.$$

At 2% end tidal halothane (equivalent to F_A = 1.88% at BTPS), the arterial pressure has halved, whereas the brain blood flow has risen to .938 ℓ/min. This gives

$$G_3(2\%) = \frac{.938}{46.5} = .0202 \text{ litres min}^{-1} \text{ mmHg}^{-1}.$$

Solving (2.12) gives K_3 = .645 litres min^{-1} mmHg^{-1}. These redistribution factors are listed for all tissues in Table 2.2. The conductance of the brain increases whereas the conductances of the other compartments decrease or remain constant with increasing arterial halothane (Smith *et al.*, 1972). The result of this is that cerebral blood flow increases then decreases with increasing depth of halothane anaesthesia as noted by Smith *et al.* (1972).

Blood pressure

Mean systemic arterial pressure is obtained from (2.13) which is the product of cardiac output and total peripheral resistance (the reciprocal of the sum of the tissue conductances). Normal blood pressure is 120/80 mmHg. Pulse pressure is proportional to stroke volume

according to Rushmer (1970)

$$P_p = 40\ Q_s' \text{ mmHg} \qquad (2.16)$$

and systolic pressure is calculated as indicated by Rushmer (1970)

$$P_s = \bar{P}_a + \frac{2}{3} P_p \text{ mmHg} . \qquad (2.17)$$

Similarly, diastolic pressure is calculated as

$$P_d = \bar{P}_a - \frac{1}{3} P_p \text{ mmHg.} \qquad (2.18)$$

2.4 INTERACTION

The steps involved in running through a simulation are listed in Table 2.3.

Table 2.3 Interaction with the program

1. Select Model (1-4)
2. Model explained with references
3. Anaesthetic circuit explained
4. Choose simulated patient's cardiac output
5. Choose simulated patient's alveolar ventilation and dead space ventilation
6. Set knobs to initial values for halothane and fresh gas flow
7. Simulation starts:

   Program ⟶ Graph produced continuously

   Student ⟶ alter gas settings
   alter graph
   alter simulation speed
   stop and restart

Setting up the simulation

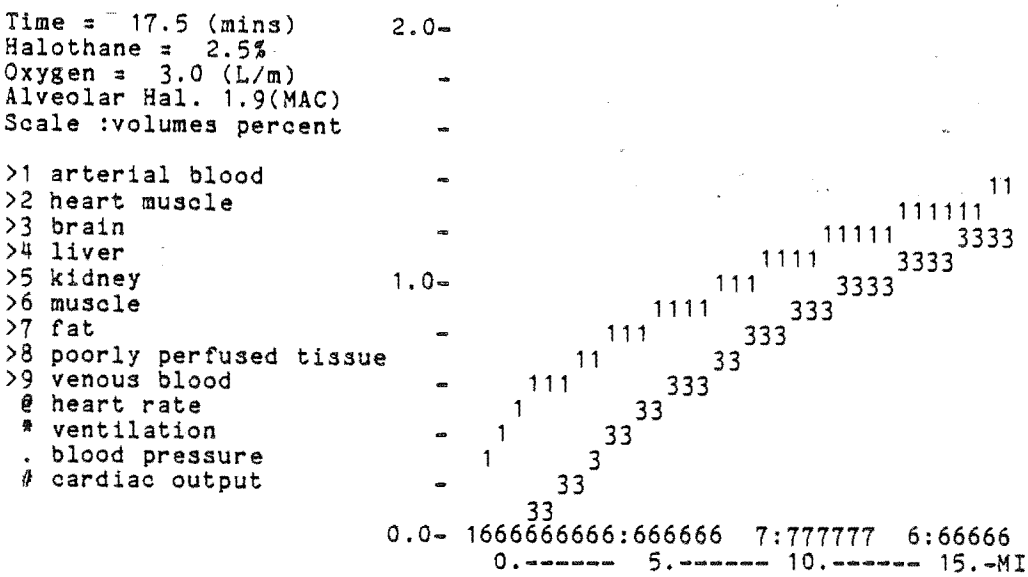
Once the student has requested this program, he selects one of four models. The simplest is a model in which ventilation, cardiac output and its distribution to the tissues remain constant. The second

introduces the effects of halothane on cardiac output and its distribution. The third model includes only the effects of halothane on ventilation and the fourth combines all of the effects outlined in the last section.

The student is then led through an introduction which describes the anaesthetic circuit and the model he has chosen (steps 2 and 3). He next selects the alveolar ventilation, dead space ventilation and cardiac output for the patient he wishes to simulate (steps 4 and 5).

Graphical output

To avoid visual confusion, the program is designed to produce graphs of selected variables on the student's terminal. Only alpha-numeric VDU terminals are available for this program so the graphs produced are formed by characters placed on the correct part of the screen (see Fig. 2.2). The resolution is sufficient for the students to clearly see the dynamics which occur.

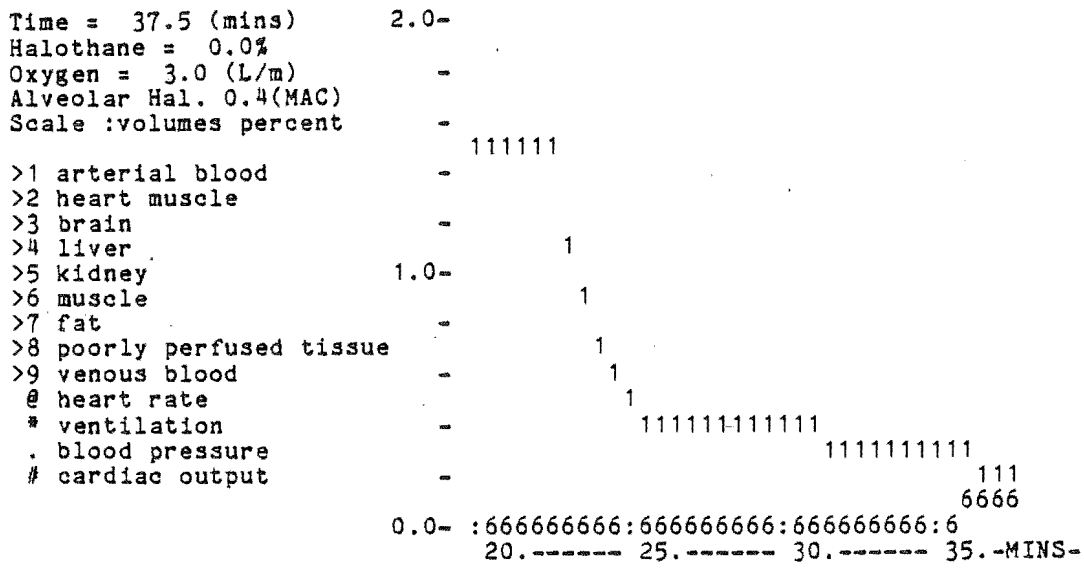


<Type 'F' to speed up, 'S' to slow down and 'H' to halt the simulation>

Figure 2.2 Terminal display for a run on the model with all effects included. Halothane is set to 2.5% from the start. For Figures 2.2 to 2.6, initial alveolar ventilation = 5 ℓ/min, dead space ventilation = 2 ℓ/min, cardiac output = 6 ℓ/min and 3 ℓ/min fresh gas (oxygen) supplied.

Down the left-hand side of the screen is a legend for the variables that can be graphed (1,2,3,...#) (see Fig. 2.2). By typing any of these symbols on the keyboard, the student causes that variable to be displayed

graphically. Typing it again causes graphing of that variable to cease. Figure 2.2 shows the effect of typing a '6' at 8 minutes and then typing it again at 14 minutes. At 9 minutes a '7' was typed to add fat to the display. This was repeated at 13 minutes to remove it. The scale on the vertical axis is automatically updated to correspond to the last variable added to the graph. The units of the scale are shown as the fifth item from the top in the left-hand margin and the variables to which the current scale applies are marked with a '>' symbol in the left-hand margin (see Figures 2.2 - 2.6).



<Type 'F' to speed up, 'S' to slow down and 'H' to halt the simulation>

Figure 2.3 Continuation of the run of Figure 2.2 with halothane turned off at 22 minutes.

Figure 2.3 is the result of removing the display of brain halothane '3' and letting the simulation of Figure 2.2 continue past the '20 minute' width of the horizontal axis. The graph has therefore 'wrapped round' to display 20 to 40 minutes on the screen.

Setting fresh gas flow and % halothane

The student is provided with a box with two knobs on it for this purpose. One controls the oxygen flow rate into the anaesthetic circuit in unit steps from 0 to 10 litres per minute ATPD. The other controls the per cent halothane that is added to this oxygen in .5% steps from 0 to 5%. The program instructs the student to set these before the simulation starts (step 6), but they can be altered at any time during the simulation.

Figure 2.3 shows the effect of setting the halothane initially at 2.5% and then dropping it to 0% after 22 minutes.

Changing the simulation speed

During the simulation the student may type an 'F' or 'S' to speed-up or slow down the simulation relative to real time. These keys can be depressed as desired, but the simulation speed is limited to eighty times real time by the computations that are done in the program. However, the program is generally run at a much slower speed, at least until the student becomes familiar with the display.

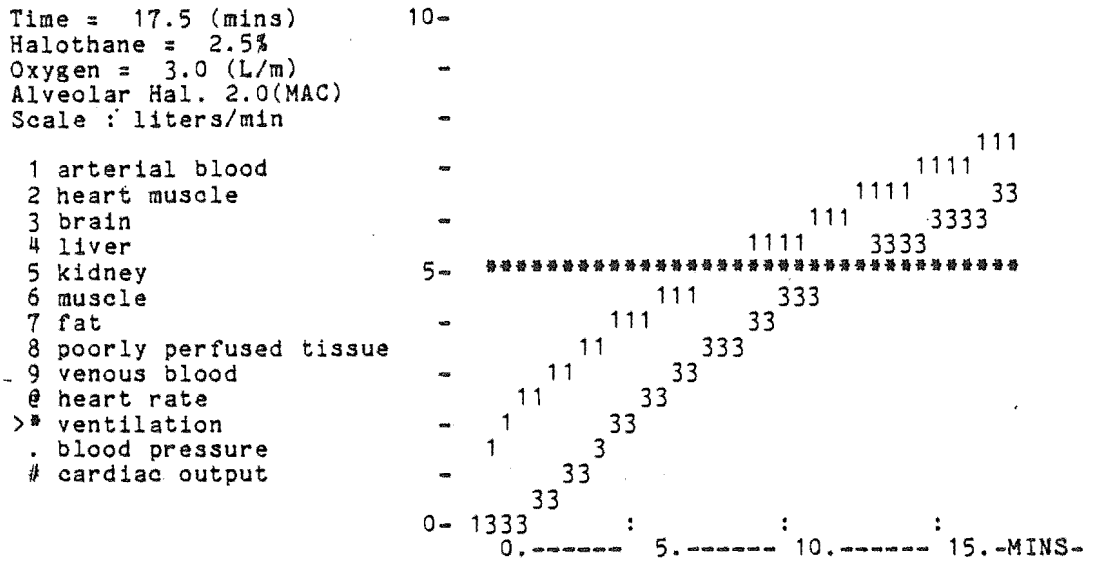
2.5 TYPICAL SIMULATIONS

Some of the more common simulations used during teaching are shown in Figures 2.2 - 2.6.

Figure 2.2 shows compartmental concentrations following a step change in inspired halothane. The arterial level - '1' (and that in the alveoli since there is no venous admixture included in the model) follows the step quickly, followed by the fast tissues, such as brain '3'. Levels in muscle - '6' and fat - '7' rise very slowly and even if the simulation were continued for another 20 minutes beyond that shown in Figure 2.2, the level of halothane in fat would not rise sufficiently to show above zero on the display.

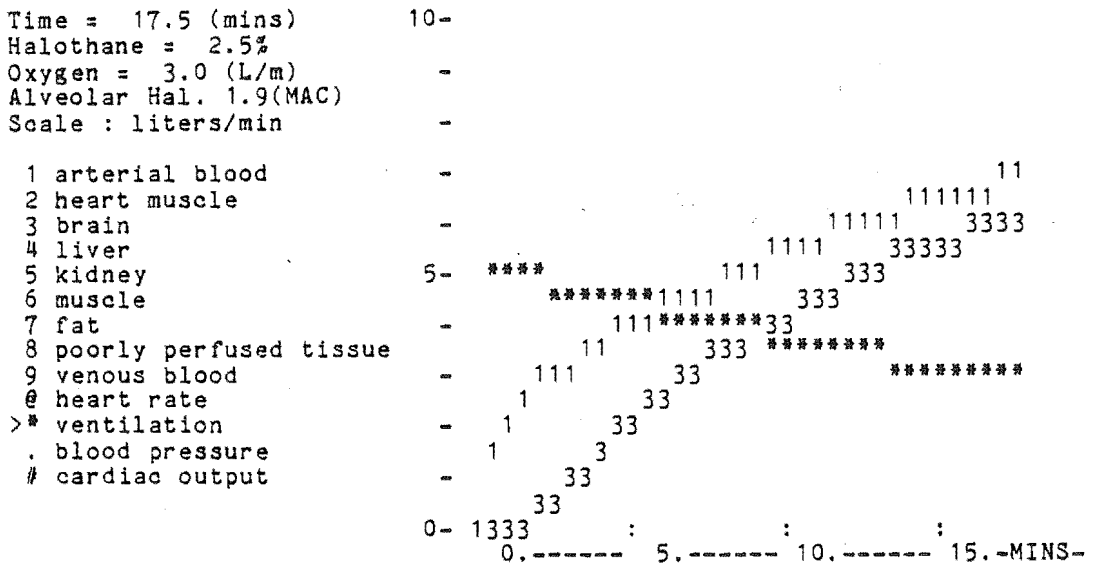
Figure 2.3 shows the decay curve for arterial concentration once the halothane is turned off. While the arterial halothane is falling it is still taken up by those tissues such as muscle - '6' whose levels are lower.

The difference between controlled ventilation (model #2) and spontaneous ventilation (model #4) is shown by the simulations in Figures 2.4 and 2.5. After 17.5 minutes of controlled ventilation at 5 litres per minute the alveolar halothane rises to 2.0 MAC (Fig. 2.4). In contrast, after breathing spontaneously for the same period the alveolar halothane reaches only 1.9 MAC (Fig. 2.5) while ventilation has fallen to 3 litres per minute.



<Type 'F' to speed up, 'S' to slow down and 'H' to halt the simulation>

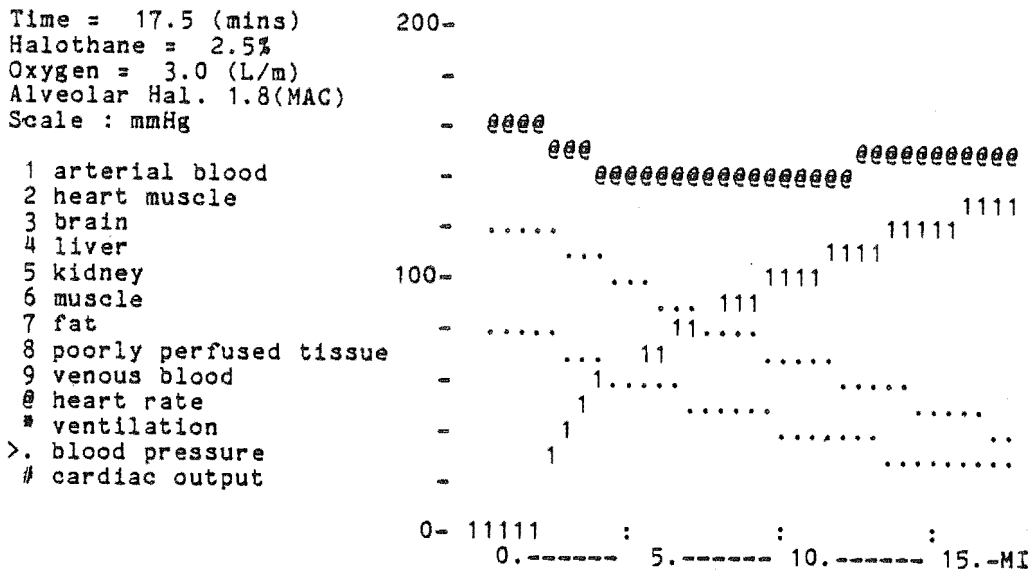
Figure 2.4 Terminal display for a run on the model which has only the cardiovascular effects included. Halothane set to 2.5% from the start.



<Type 'F' to speed up, 'S' to slow down and 'H' to halt the simulation>

Figure 2.5 Terminal display for a run using the same settings as Figure 2.4 but using the model which has all the effects included.

On induction of anaesthesia with 2.5% halothane the model predicts an initial drop of heart rate - '@' from 80 to 70 beats/minute and a subsequent slow recovery (Fig. 2.6). The blood pressure - '.' falls continuously from an initial 120/80 mmHg to 40/30 mmHg after 17 minutes by which time the arterial concentration reaches 1.8 MAC. Note that the



<Type 'F' to speed up, 'S' to slow down and 'H' to halt the simulation>

Figure 2.6 Display for a run with the same model as used in Figures 2.2, 2.3 and 2.5. Halothane was turned from 0 to 2.5% at 1.5 minutes.

scale in Figure 2.6 is for blood pressure (mmHg). Full scale for heart rate - '@' is 0-100 beats/minute.

2.6 EVALUATION

This Computer Aided Learning (CAL) program simulates the pharmacodynamics of halothane including the way it affects the circulation and ventilation of an isolated subject. The program calculates quantities (such as brain halothane levels) which cannot normally be measured and thus in common with other CAL systems (Rubin *et al.*, 1977) the student is given an experience which is difficult to provide in any other way.

To assess such programs, subjective questionnaires have been used (Rubin *et al.*, 1977), as well as comparisons between CAL groups and groups receiving traditional teaching (Kenny and Schmulian, 1979). In order to remove the teacher and student bias from an evaluation it is necessary to consider large numbers of students and different tutors. This approach was not possible here as there was only one tutor and 23 students available, each of whom had time only for a 1 hour tutorial. Thus a 'before and after' style test was felt to be as objective as could be achieved.

The evaluation was carried out by having the students complete a multichoice questionnaire before and after a tutorial session with the program conducted by a lecturer. The tutorial session in front of the computer terminal lasted approximately one hour and the students were taken in groups of two or three although each student filled out the questionnaire independently.

During the tutorial, simulations were demonstrated to the students as follows:

1. Blood and brain levels of halothane were graphed for various percentages of inspired halothane, using a fresh gas flow of 3 litres per minute (as in all these examples), over a period of 20 minutes. This illustrated the relationship of rate of change of brain and arterial levels and the concept of 'gradients' between compartments.
2. Number 1 was repeated, but with other tissue levels displayed to show the importance of tissue volume, partition coefficient, and blood flow in determining the relative rates of change of halothane levels.
3. The halothane was turned off after administration at a constant rate for 1 hour. Examination of the relative rates of fall in level in the various tissues showed redistribution of halothane from brain to muscle and fat.
4. Halothane was administered at a constant rate until the brain concentration reached equilibrium at a level corresponding to 1.5 MAC. By observing the graphs of brain halothane levels the times for equilibrium to be reached were compared between models in which ventilation and/or cardiac output were held constant. This illustrated the effect of halothane on ventilation and cardiac output and in particular the limits placed on tissue level changes during spontaneous ventilation in comparison with those during controlled ventilation.
5. Heart rate, blood pressure and ventilation, and the halothane level in the brain, were graphed during an exercise in which anaesthesia was induced with a high initial inspired halothane, which was gradually reduced to maintain MAC at 1.5%. This allowed the student to examine the 'clinical' effects of a prescribed MAC, to see how this could be rapidly attained and how the effects persisted after 'anaesthesia' of varying duration.

The questionnaire is listed in Appendix 1. Questions 1-4 were given to the students before the tutorial session and were repeated at the end as questions 5-8. This 'before and after' sequence was intended to see if the students retained information presented in the tutorial session. The significance of the changes in the students' replies was determined by the McNemar test for the significance of changes (Siegel, 1956). In the cases where there were insufficient numbers to apply this test (less than 5), the significance was assessed by the Binomial test (Siegel, 1956).

Questions 9 and 10 were intended to see if the students understood the form of the display on the computer terminal, whereas question 11 was to see if the students understood the limitations of this computer simulation. Questions 13 and 14 asked whether the students thought the program was worthwhile. Space was left at the end of the questionnaire for general comments.

The questionnaires were given to a total of 23 students.

Table 2.4 summarises the improvement in the students' responses from before (1-4) to after (5-8) the tutorial. The McNemar test for significance of changes (Siegel, 1956) was applied to the pooled results and was found to be highly significant ($P < .00001$).

Table 2.4 Summary of the results for the first eight questions.

Question	Wrong /wrong	Wrong /right	Right /wrong	Right /right	Total	Improvement*
1 and 5	3	15	0	3	21	83%
2 and 6	5	8	1	4	18	61%
3 and 7	11	6	1	0	18	35%
4 and 8	4	7	3	7	21	63%

* This column is the students who improved (wrong/right), given as a percentage of those who initially got this question wrong.

The results are not so significant if the questions are considered separately. By the same test, question 1 and 5 gives $P < .0003$ and question 4 and 8 gives $P < .34$ which is not significant. The Binomial test (Siegel, 1956) gives $P < .02$ for question 2 and 6 and $P < .109$ for question 3 and 7.

All the students replied correctly to question 9, but seven out of the 23 answers for question 10 were incorrect.

From 22 replies to question 11, 16 picked the correct combination of 1 and 4 and a further four selected option 4 as one of their answers. No one chose options 2 or 3.

Question 12 determined that only one student had used the program before.

For question 13, 18 out of 22 gave answer #1 and the other four gave option #2. Nineteen students gave #1 as a reply to question 14 and seven gave #2 (some students gave multiple replies to this question).

Typical comments given at the end of the questionnaire were:

'Would have liked to be introduced to the computer earlier in the year.'
'It is difficult to follow the use of it.'
'Helpful teaching method. Would need a couple of sessions like this with advisor/staff member present.'
'Good teaching method.'

2.7 DISCUSSION

The program described here has been designed to simulate the pharmacokinetics of halothane administration together with some of the modifications of physiological interactions which occur when the drug is administered to a subject. In order to facilitate the teaching of this material the subject is removed from the complexities of an actual anaesthetic.

To enable the importance of the various physiological effects of halothane to be studied in isolation, the student is given the choice of four models of varying complexity. The first of these allows the student to see uptake and distribution of halothane apart from its influence on the respiratory and cardiovascular systems. In the second model, halothane affects the cardiovascular system only whereas in the third it affects only ventilation. The fourth combines both effects.

For example, by comparing results from the second model with the fourth (as in Figs 2.4 and 2.5) the student can see the difference between controlled and spontaneous ventilation. Clearly, the depression of ventilation in the spontaneously breathing patient acts as a safeguard,

and limits further uptake of halothane (as noted by Munson *et al.*, 1973).

The model described here incorporates the best available data and gives qualitatively acceptable predictions (Figs 2.2 - 2.6) in terms of halothane levels and respiratory and cardiovascular values. It combines the effects of halothane on blood flow and ventilation as described separately by others (Deutsch *et al.*, 1962; Eger *et al.*, 1971; Smith *et al.*, 1972; Munson *et al.*, 1973; Eger, 1974). Independent validation of the model (cf. Dickinson, 1977) is yet to be carried out.

The application described in this chapter is an example from the third class of CAL program described in §1.4. Such programs suffer from the difficulties associated both with the development of a model and with the design of interaction. The model, which is described in §2.3, was developed using data published in the literature. This process was not straightforward, however, since there is little consistent data available on certain aspects of halothane anaesthesia. In such situations the modelling process was guided by the clinical experience of several members of the staff of the Department of Anaesthesia, Christchurch Clinical School of Medicine. The interface between the computer and the student is a very important aspect of any CAL program (cf. §1.4) and this has been made as simple and clear as possible. The simulation makes full use of the terminal for graphics display within the limits of an alphanumeric VDU. While the program is running the student can:

1. Alter halothane and fresh gas flow.
2. View continuous graphs on the terminal which 'wrap around' as the 20 minute screen 'width' is exceeded. 'Anaesthesia' can be as long or as short as desired.
3. Choose and alter the information displayed on the screen during a simulation.
4. Alter simulation speed relative to real time.
5. Choose the next action by viewing the prompts displayed at the side of the screen.

The overall results from the first eight questions of the evaluation questionnaire suggest that information presented in the tutorial was retained by the students at least for a short time. McNemar's test demonstrates that the tutorial session helped rather than hindered learning by comparing the number who improved (wrong/right) with the number who learned incorrect knowledge (right/wrong). It takes

no account of those whose level of knowledge stayed the same. Therefore, an important measure of the utility of the CAL tutorial session is the proportion of students who initially answered a question wrongly, but who gained knowledge from the tutorial. The figures listed in Table 2.4 clearly show that most of the students initially wrong gained knowledge except in question 3 where only one third showed improvement.

In attempting to evaluate a computer program used as a teaching aid, it is clearly difficult to separate the learning engendered by the program from that attributable to the tutor. Such a separation has not been achieved in the present study. Kenny and Schmulian (1979) compared CAL and tutorial teaching, finding the former significantly better in terms of a "post teaching" test. It can only be said that the combination of tutor and computer program produces a significant increase of short-term knowledge of the uptake and distribution of halothane. As Dickinson (1977) says, it is enough if the student's interest is aroused, provided knowledge or habits that are wrong or undesirable are not being taught. Questions 9, 10 and 11 show that the students understand the purpose of this form of presentation and questions 13 and 14 indicate that it raised their interest.

The great strength of the computer simulation is that it stresses principles that cannot easily be demonstrated during clinical anaesthesia. The presence of the tutor ensures that the students do not 'gloss over' simulation results which they may not fully understand.

Modern technology has made it possible to simplify the interaction at a computer terminal. This is exploited to the full in the present program and the response to question 14 indicates that the students feel the program is sufficiently flexible and easy to use so that they are attracted to use it by themselves.

The program has also been used with groups of registrars training in anaesthesia. These students have practical experience of halothane anaesthesia and some of them have been attracted to use the program to study the effects that they have observed in the theatre.

3. BLURRING DUE TO OBJECT MOVEMENT

3.1 STATEMENT OF THE PROBLEM

A brief review of the effect of object movement during CT scanning has already been given in §1.6. The approach taken in this chapter is to examine the loss of resolution caused by different types of object movement.

In practice, the image plane is a particular cross-section of a body (e.g. the cranium of a hospital patient, etc.). When the body moves, the distribution of density over the image plane varies with ϕ. It makes sense to re-write $\lambda(r;\theta)$ as $\lambda(r;\theta;\phi)$, so that the projection at angle ϕ is defined by

$$f(\xi;\phi) = \int_{-\infty}^{\infty} \lambda(r;\theta;\phi)\, d\eta \qquad (3.1)$$

rather than by (1.3).

If the form of the body movement is known *a priori*, it can be allowed for, provided that the whole body moves rigidly in the image plane (cf. Kowalski and Wagner, 1977). What cannot be allowed for are arbitrary movements of parts of the body with respect to other parts. It is useful to introduce an estimated reconstructed image $\hat{\lambda}(r;\theta)$ defined by substituting (1.4) into (1.5). Thus

$$\hat{\lambda}(r;\theta) = \int_0^{2\pi} \int_0^{\infty} \int_{-\infty}^{\infty} f(\xi;\phi)\, \exp\big(i2\pi\rho[\xi - r\cos(\theta-\phi)]\big)\, \rho\, d\xi d\rho d\phi \qquad (3.2)$$

where $f(\xi;\phi)$ is given by (3.1).

It is clear that the effect of any sort of body movement in the image plane can be built up from projections of distributions of points that move arbitrarily in the plane. The effect of a distribution of points can be built up from the effect of a single point having a density that is zero everywhere except at, say, the point Q of Figure 3.1 where the density is infinite (but the "mass" is unity). Consequently it is sufficient to set

$$\lambda(r;\theta,\phi) = \delta(x - \tau\cos\psi)\, \delta(y - \tau\sin\psi) = \delta(r-\tau)\, \delta(\theta-\psi)/r \qquad (3.3)$$

where $\delta(\cdot)$ denotes the Dirac delta function, and where $\tau = \tau(\phi)$ and

$\psi = \psi(\phi)$. It follows that

$$f(\xi;\phi) = \delta(\xi - \tau\cos(\psi - \phi)) \tag{3.4}$$

which further implies that

$$\hat{\lambda}(r;\theta) = \int_0^{2\pi} \int_0^{\infty} \exp(i2\pi\rho[\tau\cos(\psi - \phi) - r\cos(\theta - \phi)])\ \rho\ d\rho\ d\phi. \tag{3.5}$$

Note that

$$\int_0^{2\pi} \int_0^{\infty} \hat{\lambda}(r;\theta)\ r\ dr\ d\theta = 1 \tag{3.6}$$

follows immediately from (3.5) because

$$\int_0^{2\pi} \int_0^{\infty} \exp(i2\pi\rho[\tau\cos(\psi - \phi) - r\cos(\theta - \phi)])\ r\ dr\ d\theta = \delta(\rho)\ \delta(\phi)/\rho. \tag{3.7}$$

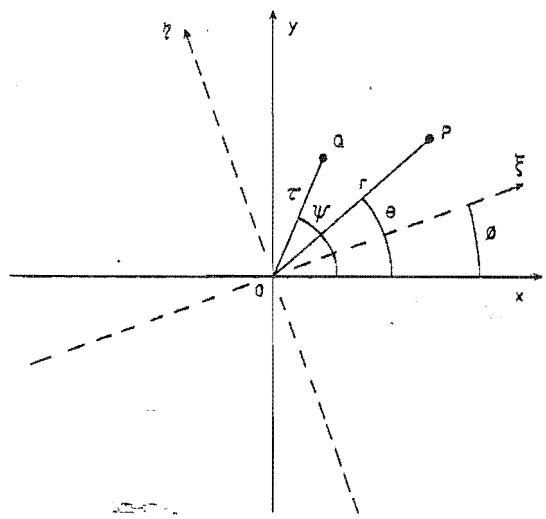


Figure 3.1 Coordinate systems in the Image Plane.

The definition (3.1) of a projection implies that it is measured by a beam of rays, each of which is parallel to the η-axis of Figure 3.1. In much computed tomography practice, projection data are obtained with fan beams rather than parallel beams (cf. §1.6). The fan beam data are processed directly in some cases, but in other cases they are "rebinned" after part or all of the data are gathered so that the image reconstruction algorithms can be based on the established theory for parallel beam projections (cf. Peters and Lewitt, 1977).

An attractive feature of the approach taken here to estimate the effects of object movement is that it is independent of how the projection data are gathered and whether or not they are rebinned before being processed. This comes about because only one ray passes through the point Q of Figure 3.1 when a single projection is measured, whether a parallel beam or a fan beam is employed. The functional forms of $\tau(\phi)$ and $\psi(\phi)$ depend upon whether or not the data are rebinned, but this difference disappears when statistical averages are taken for classes of object movement.

Figure 3.2a is a relief plot of a reconstructed density $\hat{\lambda}(r;\theta)$.

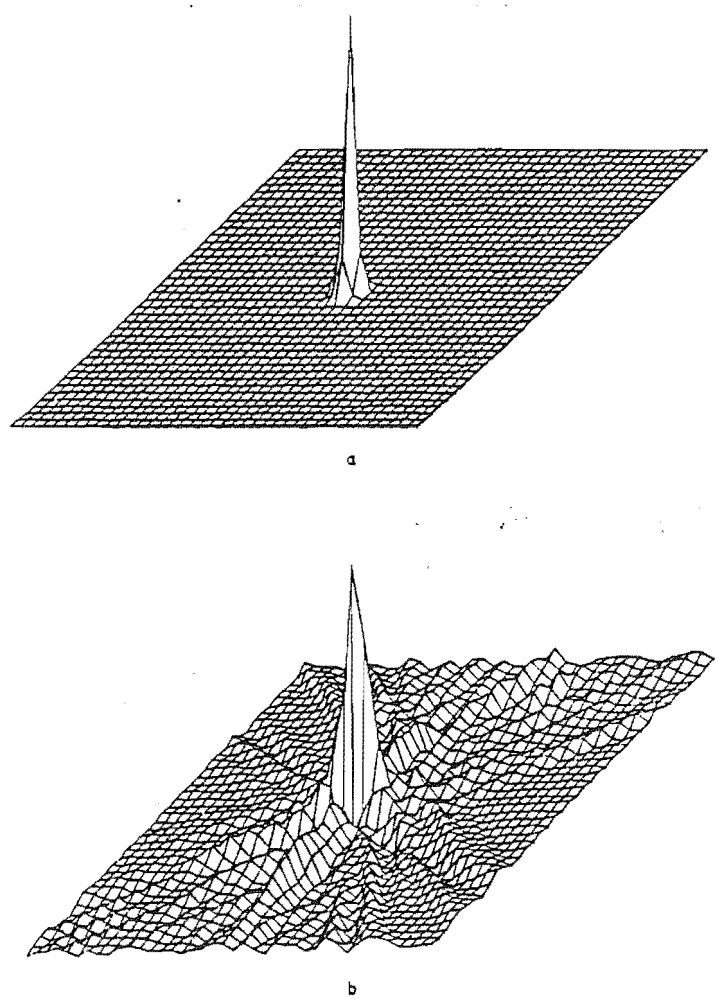


Figure 3.2 Relief maps of images reconstructed from 96 projections each of 120 points, modified by the Shepp-Logan filter (cf. §1.6) and back-projected on to a 120×120 grid. A 40×40 central section of this grid is shown.
(a) Q permanently coincident with 0.
(b) Q coincident with 0 for 33% of the time; Q wanders at random along x-axis for the rest of the time.

The projections used to reconstruct this image were calculated from an object which consisted of a delta function at the point Q which was coincident with the point O while all the projections were 'measured'. The image exhibits little blurring, and is a good representation of the original object. Figure 3.2b shows the form of the reconstruction when Q wanders back and forth along the x-axis while the projections are 'measured' (see figure caption for details). The image of the delta function is now badly blurred.

The numerical evaluation of (3.5) (e.g. by modified back-projection) required to produce Figure 3.2b is much too protracted to form the basis of a computationally efficient means of evaluating blurring by various kinds of motion. It is also too complicated. A simple description of the blurring is required. Consequently, a theoretical device is introduced in §3.2 to permit the spread of $\hat{\lambda}(r;\theta)$ to be estimated quasi-statistically. Archetypal movements are defined in §3.3. Computational results for the blurring caused by these movements are presented in §§3.4 and 3.5. The technical significance of the results is assessed in §3.6.

3.2 TEST FUNCTION

To satisfy the requirements introduced in the final paragraph of §3.1, a single parameter, descriptive of the degree of blurring, needs to be obtained from (3.5). There are an infinite number of possible movements, but any particular movement can be characterised by the fraction of time that Q departs by more than a given distance from its central position. It is convenient to consider that this central position is coincident with O. This does not restrict the types of movement that are studied here since the origin can notionally be shifted to any point in the plane. It is only a slight restriction to further require that the form of any motion must be symmetrical about its central position (defined to be the origin). This means that

$$\int_0^{2\pi} (\tau \cos(\psi - \phi))^{2n+1} \, d\phi = 0 \qquad (3.8)$$

where n is any integer. Equation (3.8) simplifies later analysis apparently without significant effect on the validity of the final conclusions. The approach taken is to multiply $\hat{\lambda}(r;\theta)$ by a smooth, peaked function having a radius σ, say. The fraction of $\hat{\lambda}(r;\theta)$ which

lies within a radial distance of σ from O characterises the spread of $\lambda(r;\theta)$. Consider the *test function* $T(\sigma)$ which is defined by

$$T(\sigma) = \int_0^{2\pi} \int_0^{\infty} \hat{\lambda}(r;\theta) \exp(-r^2/(2\sigma^2)) \, r\,dr\,d\theta \, . \tag{3.9}$$

$T(\sigma)$ is used to represent the reconstructed density within a radius σ. If σ is much larger than the effective radial spread of $\hat{\lambda}(r;\theta)$ then (3.6) shows that $T(\sigma)$ tends to unity. Since the object is a delta function, the contribution at the origin ($\sigma = 0$) gives the fraction of the reconstructed density which is not blurred. If there is no movement ($\tau = 0$) the reconstructed density is given by

$$\hat{\lambda}(r;\theta) = \lambda(r;\theta) = \delta(r)\,\delta(\theta) / r \, . \tag{3.10}$$

Substituting (3.10) into (3.9) shows that

$$T(\sigma) \equiv 1 \tag{3.11}$$

implying, of course, that there is no blurring. However, if there is movement, the amplitude of the test function decreases rapidly once the value of σ falls below the spread of the movement, implying that there is some blurring.

Substituting (3.5) into (3.9) shows that

$$T(\sigma) = \int_0^{2\pi}\int_0^{2\pi}\int_0^{\infty}\int_0^{\infty} \exp[i2\pi\rho\tau\cos(\psi-\phi) - i2\pi\rho r\cos(\theta-\phi) - r^2/(2\sigma^2)]\,\rho r \, d\rho\, dr\, d\theta\, d\phi. \tag{3.12}$$

Equation 9.1.41 of Abramowitz and Stegun (1965) shows that

$$\exp[-i2\pi\rho r\cos(\theta-\phi)] = \sum_{k=-\infty}^{\infty} (-i)^k \exp[ik(\theta-\phi)]\, J_k(2\pi\rho r) \tag{3.13}$$

where $J_k(\cdot)$ denotes the Bessel function of the first kind of order k. The θ integration now yields

$$T(\sigma) = 2\pi \int_0^{2\pi}\int_0^{\infty}\int_0^{\infty} \exp[i2\pi\rho\tau\cos(\psi-\phi) - r^2/(2\sigma^2)]\, J_0(2\pi\rho r)\, \rho r \, d\rho\, dr\, d\phi. \tag{3.14}$$

Weber's first exponential integral (cf. Watson, 1966, equation 1 of §13.3) enables the r-integral to be done, giving

$$T(\sigma) = \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\infty} \exp[iR\tau\cos(\psi-\phi)/\sigma - R^2/2]\, R\, dR\, d\phi \, , \tag{3.15}$$

where $R = 2\pi\sigma\rho$.

This is now in a form suitable for combined analytical and numerical evaluation, because the magnitude of the integrand, unlike that in (3.5), falls off rapidly with increasing radial coordinate, when the latter is large enough.

3.3 ARCHETYPAL MOVEMENTS

The strategy adopted here is to analyse sufficient species of movement to be able to reach general conclusions concerning the degree of blurring to be expected from body movement. Two *classes* of movement are studied:

A. *Continuous movements* for which $\tau(\phi)$ and $\psi(\phi)$ are everywhere differentiable with respect to ϕ.

B. *Discrete movements* for which $\tau(\phi)$ and $\psi(\phi)$ are continuous for finite ranges of ϕ, but jump discontinuously when ϕ leaves one range and enters the next.

Two different *types* of movement are discussed in §§3.4 and 3.5. For each type, particular movements of *Classes* A and B are investigated.

Because $f(\xi;\phi+\pi) = f(-\xi;\phi)$, as noted in §3.1, it follows that τ and ψ must satisfy

$$\tau(\phi + \pi) = \tau(\phi) \qquad \text{and} \qquad \psi(\phi + \pi) = \psi(\phi) \; . \tag{3.16}$$

A convenient description of the "amount of movement" is the *effective radius*, R_e, defined by

$$R_e^2 = 1/\pi \int_o^\pi \tau^2 \, d\phi \; . \tag{3.17}$$

Note that R_e is a "root mean square" quantity, and so accords with conventional notions of the "effective amplitude" of a process which is itself of zero mean. All graphs of $T(\sigma)$ presented in this chapter are plotted versus σ/R_e.

The first step in evaluating the integral in (3.15) is to expand $\exp[i(\tau/\sigma)R\cos(\psi-\phi)]$ in its absolutely convergent Taylor series, so that

$$T(\sigma) = \frac{1}{2\pi} \sum_{n=0}^{\infty} (i/\sigma)^{2n} \frac{1}{(2n)!} \int_o^{2\pi}\int_o^{\infty} R^{2n}(\tau\cos(\psi-\phi))^{2n} \exp(-R^2/2)R \, dR \, d\phi \tag{3.18}$$

where only the even terms are retained because of (3.8). The R integration is solved by parts to give

$$\int_o^\infty R^{2n} \exp(-R^2/2)\, RdR = 2n \int_o^\infty R^{2(n-1)} \exp(-R^2/2)\, RdR = 2^n\, n! \ . \qquad (3.19)$$

So (3.18) becomes

$$T(\sigma) = \frac{1}{2\pi} \sum_{n=o}^{\infty} (-2/\sigma^2)^n \frac{n!}{(2n)!} \int_o^{2\pi} (\tau\cos(\psi - \phi))^{2n}\, d\phi \ , \qquad (3.20)$$

which is the basis for some of the numerical examples presented in this chapter.

3.4 CIRCULAR MOVEMENTS

For the type of movement considered in this section, only ψ varies within each range of ϕ for which τ and ψ are continuous.

Class A movements

In order to obey (3.8) a point moving continuously in a circle about its average point is described by the parameters $\tau = R_e$, $\psi = 2m\phi$ where m is a non-negative integer. It follows from (3.15) that

$$T(\sigma) = \frac{1}{2\pi} \int_o^{2\pi} \int_o^\infty \exp(-R^2/2 + iRR_e\cos[(2m-1)\phi]/\sigma)\, RdR \ . \qquad (3.21)$$

Again equation 9.1.41 of Abramowitz and Stegun (1965) gives

$$T(\sigma) = \frac{1}{2\pi} \int_o^{2\pi} \int_o^\infty \exp(-R^2/2) \sum_{\ell=-\infty}^{\infty} (i)^\ell \exp[i(2m-1)\ell\phi]\ J_\ell(R\,R_e/\sigma)\, RdRd\phi. \qquad (3.22)$$

Doing the ϕ-integration gives

$$T(\sigma) = \int_o^\infty \exp(-R^2/2)\, J_o(R\,R_e/\sigma)\, RdR \ , \qquad (3.23)$$

which is Weber's first exponential integral (cf. Watson, 1966, equation 1 of §13.3). Thus

$$T(\sigma) = \exp\left(-R_e^2/(2\sigma^2)\right) \qquad (3.24)$$

which is, most conveniently, independent of m. Figure 3.3 shows $T(\sigma)$ versus σ/R_e.

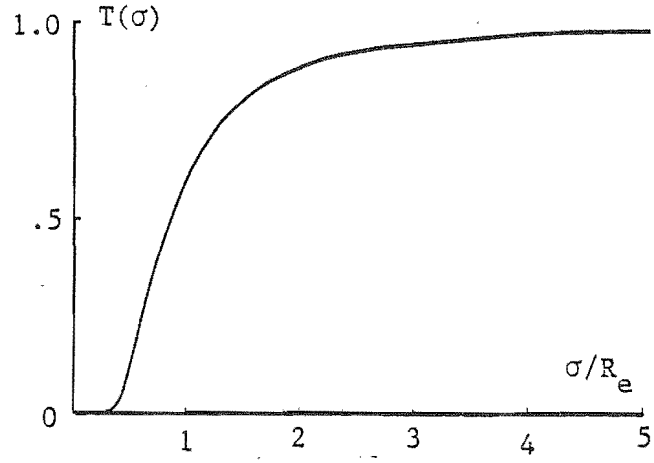


Figure 3.3 Test function for *Class A* circular movement.

Class B movements

$$\text{for} \begin{cases} 0 < \phi < \alpha : & \tau = R_e' \quad \text{and} \quad \psi = \beta + \phi s \\ \alpha < \phi < \pi - \alpha : & \tau = 0 \\ \pi - \alpha < \phi < \pi : & \tau = R_e' \quad \text{and} \quad \psi = \beta + \pi + \alpha s + (\phi - \pi)s \end{cases}$$

where $R_e' = (\pi/2\alpha)^{\frac{1}{2}} R_e$. As a result of equation (3.16)

$$\text{for} \begin{cases} \pi < \phi < \pi + \alpha : & \tau = R_e' \quad \text{and} \quad \psi = \beta + (\phi - \pi)s \\ \pi + \alpha < \phi < 2\pi - \alpha : & \tau = 0 \\ 2\pi - \alpha < \phi < 2\pi : & \tau = R_e' \quad \text{and} \quad \psi = \beta + \pi + \alpha s + (\phi - 2\pi)s. \end{cases}$$

The test function for this type of movement is calculated from (3.20). The ϕ integration becomes:

$$\begin{aligned} \int_0^{2\pi} (\tau\cos(\psi - \phi))^{2n} d\phi \;=\; & \int_0^{\alpha} (R_e')^{2n} \left(\cos(\beta + \phi(s-1))\right)^{2n} d\phi \\ & + \int_{\pi-\alpha}^{\pi} (R_e')^{2n} \left(\cos(\beta + \pi + \alpha s + (\phi - \pi)s - \phi)\right)^{2n} d\phi \\ & + \int_{\pi}^{\pi+\alpha} (R_e')^{2n} \left(\cos(\beta + (\phi - \pi)s - \phi)\right)^{2n} d\phi \\ & + \int_{2\pi-\alpha}^{2\pi} (R_e')^{2n} \left(\cos(\beta + \pi + \alpha s + (\phi - 2\pi)s - \phi)\right)^{2n} d\phi. \end{aligned} \tag{3.25}$$

The first and third, second and fourth terms can be combined to give

$$\int_0^{2\pi} (\tau\cos(\psi-\phi))^{2n}\, d\phi = \frac{2(R'_e)^{2n}}{s-1} \int_{\beta}^{(s-1)\alpha+\beta} (\cos\phi)^{2n}\, d\phi + \frac{2(R'_e)^{2n}}{s-1} \int_{\alpha+\beta}^{\alpha s+\beta} (\cos\phi)^{2n}\, d\phi \,. \tag{3.26}$$

These integrals can be computed recursively by noting that (cf. Abramowitz and Stegun, 1965, equation 4.3.127):

$$\int (\cos z)^{2n}\, dz = \frac{\sin z(\cos z)^{2n-1}}{2n} + \frac{2n-1}{2n} \int (\cos z)^{2(n-1)}\, dz \,. \tag{3.27}$$

Equations (3.20), (3.26) and (3.27) specify the scheme used to calculate the test function for this type of motion. Figures 3.4 and 3.5, respectively, show the effect on $T(\sigma)$ of varying the speed s of rotation and the angle β at which the rotation starts. Note both that Q remains at 0 for a fraction k of the complete range $[0,\pi]$ of the variation of ϕ and that k is the value to which $T(\sigma)$ reduces as σ tends to zero. The above definition of *Class B* circular movements ensures that

$$k = T(0) = 1 - 2\,\alpha/\pi \,. \tag{3.28}$$

Inspection of (3.15) indicates that $T(\sigma)$ can be expressed in the form

$$T(\sigma) = k + (1-k) \int_0^{2\pi} \int_0^{\infty} \exp(-R^2/2 + i\,R R'_e \cos[\beta+\gamma(\phi)]/\sigma)\; R\,dR\; d\phi \tag{3.29}$$

where $\gamma(\phi)$ is independent of β and where the movement defined by $\tau(\phi)$ and $\psi(\phi)$ is now assumed to cover a range of ϕ equivalent to $\alpha = \pi/2$, i.e. there is now assumed to be movement occurring for the whole time projections are measured.

In the majority of practical situations the precise nature of the body movement is unknown *a priori*. What is likely to be of interest is the average blurring to be expected from a class of movement. Consequently, it makes sense to define an *effective test function* $T_e(\sigma)$, which is $T(\sigma)$ averaged over all starting angles β:

$$T_e(\sigma) = \frac{1}{2\pi} \int_0^{2\pi} T(\sigma)\, d\beta$$

$$= k + \frac{1-k}{(2\pi)^2} \int_0^{2\pi} \int_0^{\infty} \exp(-R^2/2) \int_0^{2\pi} \exp(iRR'_e \cos[\beta+\gamma(\phi)]/\sigma)\; d\beta\, R dR\, d\phi \tag{3.30}$$

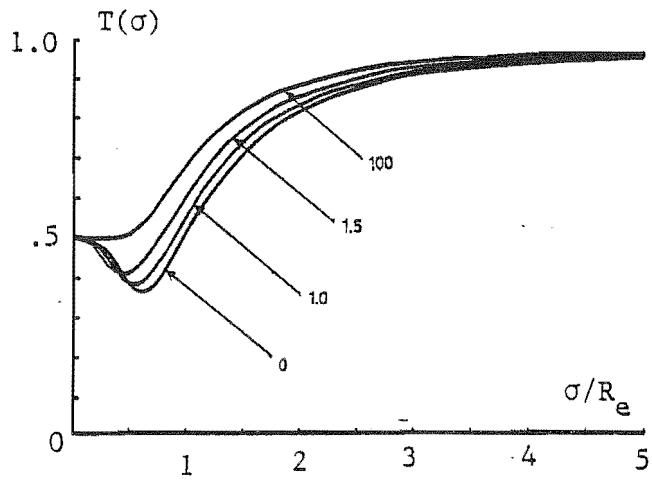


Figure 3.4 Test function for *Class B* circular movement: $\alpha = \pi/4$, $\beta = 0$ for four values of s.

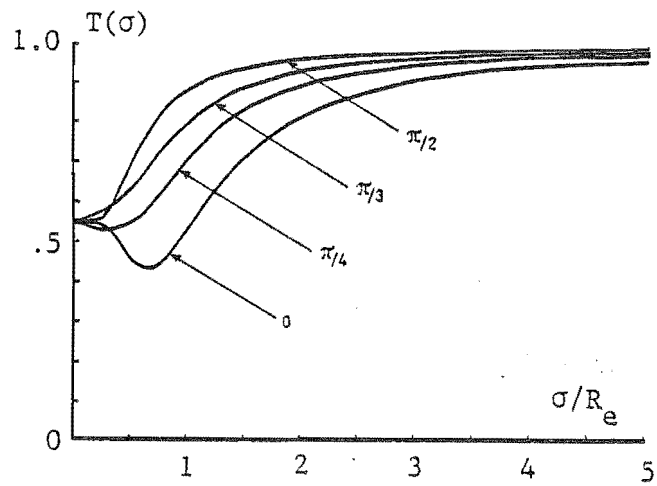


Figure 3.5 Test function for *Class B* circular movement: $\alpha = 2\pi/9$, s = 0.1 for four values of β.

The β integral is completed by substituting from (3.13) to give

$$T_e(\sigma) = k + \frac{1-k}{2\pi} \int_0^{2\pi} \int_0^{\infty} \exp(-R^2/2)\, J_0(R\, R_e'/\sigma)\, R\, dR\, d\phi \qquad (3.31)$$

which is again Weber's first exponential integral. Thus

$$T_e(\sigma) = k + (1-k) \exp\left[-(R_e')^2/(2\sigma^2)\right]. \qquad (3.32)$$

Figure 3.6 shows $T_e(\sigma)$ versus σ/R_e for three values of k.

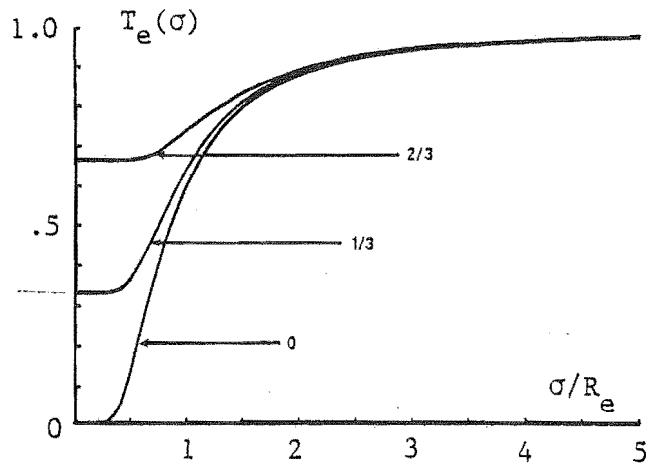


Figure 3.6 Effective test function for *Class B* circular movement: for three values of k. Note that this figure also applies to *Class B* linear movement.

3.5 LINEAR MOVEMENTS

For the type of movement considered in this section it is convenient to set $\psi = 0$. When Q wanders on to the negative x-axis τ is taken to be negative.

Class A movements

$\tau = A_\ell \sin(2\ell\phi) + A_m \sin(2m\phi)$. By allowing the constants A_ℓ and A_m (ℓ and m are integers) to successively assume a range of values, a useful variety of linear movements can be generated. The first step is to expand the integrands in (3.20) in terms of exponentials, i.e.

$$\tau(\phi)\cos(\psi-\phi) = [1-\exp(i2\phi)]\{A_m[\exp(i(2m-1)\phi)-\exp(-i(2m+1)\phi)] + A_\ell[\exp(i(2\ell-1)\phi)-\exp(-i(2\ell+1)\phi)]\}/(4i) \quad . \tag{3.33}$$

The Binomial expansion is now used so that

$$(\tau(\phi)\cos(\psi-\phi))^{2n} = \frac{1}{(-16)^n} \sum_{\xi=0}^{2n} \sum_{\alpha=0}^{2n} \sum_{\beta=0}^{2n-\alpha} \sum_{\gamma=0}^{\alpha} \binom{2n}{\xi}\binom{2n}{\alpha}\binom{2n-\alpha}{\beta}\binom{\alpha}{\gamma}(-1)^{\beta+\gamma} A_m^{2n-\alpha} A_\ell^{\alpha} \exp(i2(2m\beta+2\ell\gamma-2mn+m\alpha+n-\ell\alpha-\xi)\phi) \quad . \tag{3.34}$$

where $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$.

Thus each of the ϕ integrals in (3.20) equals zero unless

$$\xi = 2m\beta + 2\ell\gamma - 2mn + m\alpha + n - \ell\alpha \qquad \text{and} \qquad 0 \leq \xi \leq 2n \ . \tag{3.35}$$

Substituting (3.34) into (3.20) and evaluating the integrals gives

$$T(\sigma) = \sum_{n=o}^{\infty} \sum_{\alpha=o}^{2n} \sum_{\beta=o}^{2n-\alpha} \sum_{\gamma=o}^{\alpha} (-1)^{\beta+\alpha} A_m^{2n-\alpha} A_\ell^{\alpha} \left(\frac{1}{8\sigma^2}\right)^n \frac{n!\ (2n)!}{(2n-\xi)!\,\xi!\,(2n-\alpha-\beta)!\,\beta!\,(\alpha-\gamma)!\,\gamma!} \tag{3.36}$$

Equations (3.35) and (3.36) permit the test function to be calculated for Class A movements. Figure 3.7 indicates that the form of $T(\sigma)$ is comparatively insensitive to changes in A_ℓ, A_m, ℓ and m.

Class B movements

$$\text{for} \begin{cases} 0 < \phi < \alpha & ; \quad \tau = 0 \\ \alpha < \phi < \beta & ; \quad \tau = R_e' \\ \beta < \phi < \pi - \beta & ; \quad \tau = 0 \\ \pi - \beta < \phi < \pi - \alpha & ; \quad \tau = -R_e' \\ \pi - \alpha < \phi < \pi & ; \quad \tau = 0 \end{cases}$$

where $R_e' = (\pi/(2(\beta - \alpha)))^{\frac{1}{2}} R_e$. As a result of (3.16) this movement is repeated in the range $\phi = \pi$ to 2π so that

$$\text{for} \begin{cases} \pi < \phi < \pi + \alpha & ; \quad \tau = 0 \\ \pi + \alpha < \phi < \pi + \beta & ; \quad \tau = R_e' \\ \pi + \beta < \phi < 2\pi - \beta & ; \quad \tau = 0 \\ 2\pi - \beta < \phi < 2\pi - \alpha & ; \quad \tau = -R_e' \\ 2\pi - \alpha < \phi < 2\pi & ; \quad \tau = 0 \end{cases}$$

Thus (3.20) reduces to

$$T(\sigma) = \frac{1}{2\pi} \sum_{n=o}^{\infty} \left(\frac{-2}{\sigma^2}\right)^n \frac{n!}{(2n)!}\, 4\, (R_e')^{2n} \int_\alpha^\beta (\cos\phi)^{2n}\, d\phi \ . \tag{3.37}$$

which is evaluated with the aid of the recursive formula (3.27). Figure 3.8 shows the effect of varying β, keeping α constant. As in §3.4, it is noted here that what is likely to be of practical interest is the average blurring to be expected from a species of movement.

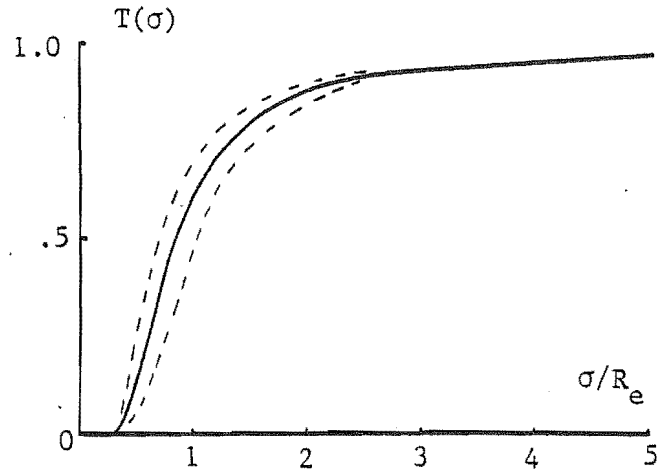


Figure 3.7 Test function for *Class A* linear movement: for ℓ and m varying from 0 to 10 and A_ℓ/A_m varying from -1.0 to 1.0. All variations of $T(\sigma)$ lie within the dashed lines.

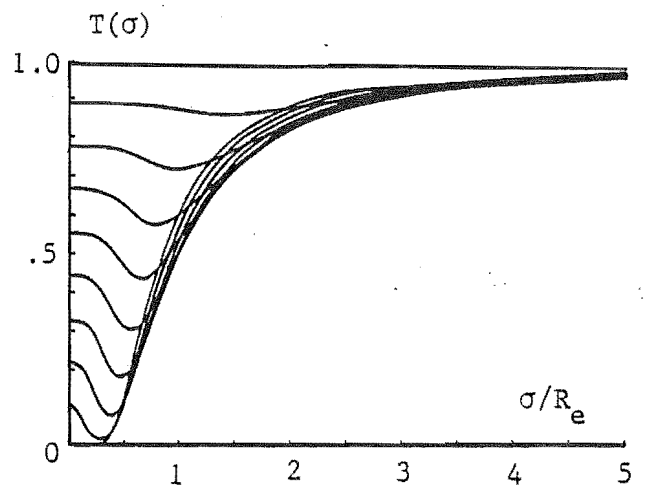


Figure 3.8 Test function for *Class B* linear movement: $\alpha = 0$, for ten values of β. For each curve: $\beta = (1 - T(0))\pi/2$.

All angles α at which the movement starts are equi-probable. By averaging over α, keeping $(\beta - \alpha)$ constant, an effective test function $T_e(\sigma)$ is obtained. $T_e(\sigma)$ is given by (3.32) with

$$k = 1 - 2\,(\beta - \alpha)/\pi. \qquad (3.38)$$

So Figure 3.6 characterises both linear and circular *Class B* movements.

3.6 DISCUSSION

The aim of this chapter has been to characterise the artefacts caused by movement. The approach taken has been to quantify the radial spread of the reconstruction of a point which moves while the projections are measured. In practice, there is never any reliable estimate of the amount of movement that has taken place during a CT scan. Thus the quasistatistical descriptor of the amount of blurring $T_e(\sigma)$ which is introduced in §3.4 is a useful measure of the loss of resolution due to motion which occurs for a fraction $(1-k)$ of the scan time. Fortunately, as was shown in §§3.4 and 3.5, the form of $T_e(\sigma)$ is independent of the form of the motion. It is only dependent on the fraction of time the point moves and the effective radius R_e' of the movement for the time it moves. Examination of (3.32) shows that it can be re-written as

$$T_e(\sigma) = k + (1-k)\, t(\sigma) \tag{3.39}$$

where

$$t(\sigma) = \exp\left[-(R_e')^2/(2\sigma^2)\right]. \tag{3.40}$$

The general descriptor, (3.39), includes the special case when movement occurs for the whole of the scan time, i.e. *Class A* movement. When a quasistatistical descriptor such as $T_e(\sigma)$ is employed to gauge the "size" of something, the effective size is often taken to correspond to the conditions applying when the descriptor is half its maximum value. Thus the effective size of the blurred image is 2σ where σ is the solution of

$$T_e(\sigma) = \cdot 5 \ . \tag{3.41}$$

For instance, when movement occurs for the whole scan time (i.e. $k = 0$), the solution of (3.41) is $\sigma \simeq .85\ R_e$. Thus the limit to which the image is resolvable is given by $1.7\ R_e$. The loss in resolution is roughly twice the effective radius of the motion.

The reconstructed image can be considered as an *attenuated blurred image* of amplitude $(1-k)$ superimposed on an *attenuated ideal image* of amplitude k. Figure 3.2b illustrates this (the caption indicates that $k = .33$). The minimum value of $t(\sigma)$ (3.40) is zero. The minimum value to which $T_e(\sigma)$ falls is given by k: the fraction of time the test point is stationary. Thus for $k > \cdot 5$, (3.41) cannot be satisfied for any value of σ. Consequently, if k is large enough that

none of the significant detail in the ideal image is submerged by the attenuated blurred image, the reconstructed image is as useful as what would have been obtained if there had been no body movement. This may explain why reconstructed images are often of high quality, even though one may feel sure that significant movement must have occurred during measurement of at least some of the projections.

Finally it is worth noting that in the classes of movement studied in this chapter the point Q has been assumed to move in the plane of the scan. In practice, there could also be movement in and out of the plane. For example in a body scan peristalsis could cause such motion. However computational experience has shown that artefacts due to so called "z-axis" motion are less severe than those due to motion in the plane (cf. Alfidi *et al.*, 1976). The results presented in this chapter can be considered to cover the "worst-case" types of motion.

4. IMAGE RECONSTRUCTION FROM FEW PROJECTIONS

4.1 INTRODUCTION

This chapter is concerned with algorithms for reconstructing an image when the number of given projections is much less than is normally needed to produce an acceptable result, i.e. projection space is undersampled in the angular direction. It is assumed, however, that the radial sampling is fully adequate for the purposes of conventional CT. Practical applications for which these conditions pertain have already been discussed in §1.6.

Consequently the concern here is with projections that are given (i.e. measured, in practice) for a comparatively small set $\{\phi_n; n=1,2...,N\}$ of values of ϕ. The given samples of $f(\xi;\phi)$ are specified for particular pairs of values of ξ and ϕ which are called the *data points*. The concise notations $\{f,\phi_n,N\}$ and $\{\phi_n,N\}$ are introduced to identify respectively, the *given projections* and the *given angles* (i.e. the values of ϕ corresponding to the given projections). This chapter is concerned with situations in which N is of the order of 10. The *interpolation problem* is posed:

Given $\{f,\phi_n,N\}$, estimate $f(\xi;\phi)$ for sufficient values of ϕ, inbetween the ϕ_n, to permit the reconstruction of $\lambda(r;\theta)$ to be acceptable, according to whatever criterion is appropriate in the envisaged application.

Not all of the algorithms presented in this chapter are explicit interpolation schemes, however. As an aid to comparison, each method is identified by a Roman numeral. Eight methods (numbered I through VIII) are compared.

Figure 4.1 shows a test phantom which is made use of in this chapter. This phantom contains sufficient structure to permit reconstruction methods to be evaluated realistically. The sharp changes in density at the edges of the various circles constitute severe tests of any image reconstruction procedure. It is difficult to assess artefacts quantitatively from inspection of conventional grey-scale images or even from relief maps. It is more convenient to examine cuts along straight lines through the images. The cuts along the lines X-X and Y-Y in Figure 4.1 are chosen for evaluating the reconstruction algorithms

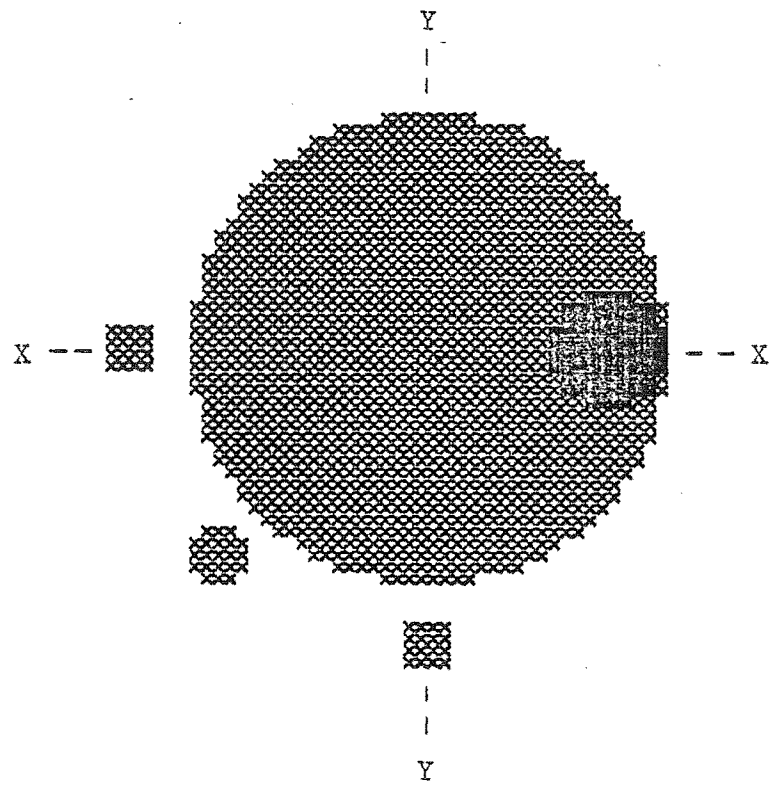


Figure 4.1 The test phantom. The dark circle has a density of 100 (arbitrary) units. The four light circles (one large and three small) are of density 50 units. The rest of the phantom is of zero density.

presented in this chapter.

A conventional root-mean-square error criterion is adopted for the 'goodness' of any of the reconstruction procedures. The *total error* $E(T,\Omega)$ is defined by

$$E^2(T,\Omega) = \iint_{\Omega} [\hat{\lambda}(r;\theta) - \lambda(r;\theta)]^2 \, d\Omega \qquad (4.1)$$

where T represents the roman numeral identifying the reconstruction method and $\hat{\lambda}(r;\theta)$ is the estimated density provided by the method. The *error criterion* is defined as

$$\varepsilon(T,\Omega) = E(T,\Omega) \; / \; E(I,\Omega) \qquad (4.2)$$

which compares any of the methods with method I. The error criterion is similar to the "distance function" of Herman *et al.* (1973). In this chapter the error criteria are calculated for two different regions of image space. The first ($\Omega = \Xi$) covers the whole of image space, while the second ($\Omega = \Lambda$) denotes the convex region which just encloses the cross-section of the object.

4.1.1 Straightforward reconstruction

The first of the reconstruction methods does not actually involve interpolation. This method (I) is reconstruction by modified back-projection using a Shepp-Logan filter (cf. §1.6.1). Figure 4.2 shows the result of applying method I when only eight projections of the test phantom are given. This image is seen to be a seriously degraded version of Figure 4.1. It is easier to see the artefacts caused by reconstruction from only a small set of projections by examining the density cuts shown in Figure 4.3.

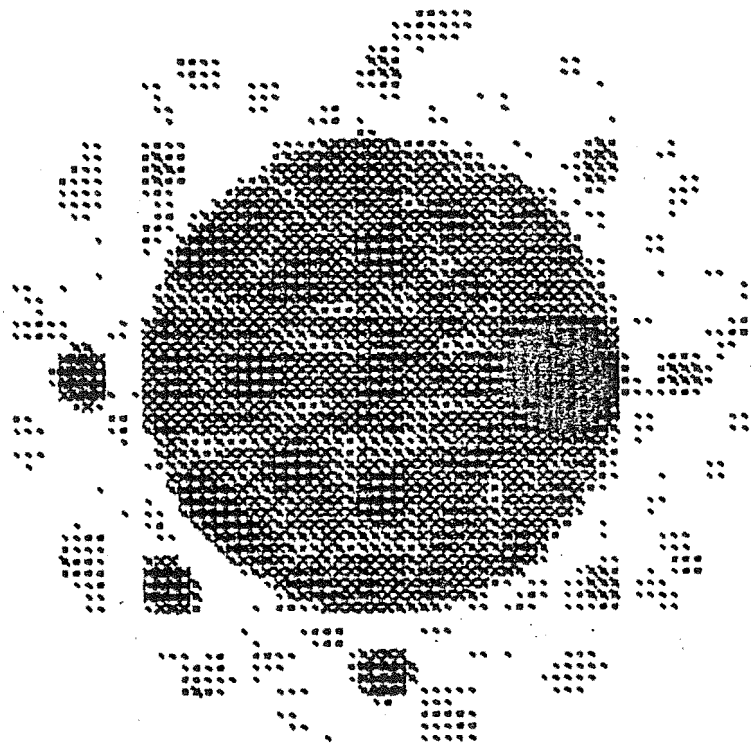


Figure 4.2 Reconstruction of the text phantom using method I (modified back-projection) when there are eight given projections. The grey-scale covers the range of zero (white) to 100 (black).

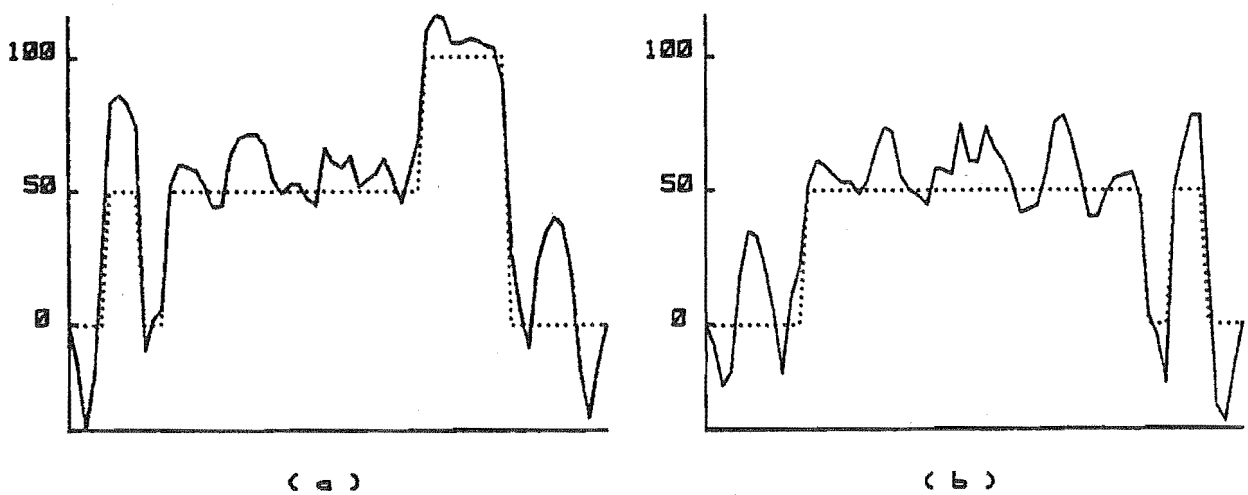


Figure 4.3 Density cuts through the test phantom (•••) and its reconstruction by method I (——): (a) cut X-X, (b) cut Y-Y.

The image suffers from angular aliasing and clutter (cf. §1.6.1). Table 4.1 lists the error criteria for method I for different values of N.

Table 4.1 Comparison criteria for method I using different numbers of given projections (N), applied to the test phantom. The error criterion $\varepsilon(T,\Omega)$ as defined by (4.2), is listed for $\Omega = \Xi$ and $\Omega = \Lambda$ (the former applies to the whole of image space, while the latter denotes the convex region actually occupied by the cross-section of the object). The error criteria are expressed as a fraction of the error criteria for the reconstruction from eight given projections.

N	$\varepsilon(I,\Xi)$	$\varepsilon(I,\Lambda)$
8	1.	1.
10	.75	.99
12	.65	.85
14	.55	.77
16	.50	.66
32	.32	.53
64	.30	.51

Table 4.1 serves as a standard of comparison for the reconstruction methods II through VIII introduced later in this chapter. Note that $\varepsilon(I,\Xi)$ decreases more rapidly than $\varepsilon(I,\Lambda)$ as the number of projections is increased. This shows the value of calculating the error criteria for different regions of image space. The second region Λ is included in the first Ξ, which emphasises that the artefact level drops off faster outside the region Λ within which the actual image exists.

4.1.2 Fourier series interpolation and straightforward reconstruction

Fourier Series Interpolation (FSI) is the second of the interpolation schemes investigated here, i.e. it is called method II. *Intermediate projections* $\hat{f}(\xi;\phi)$ are computed, for values of ϕ not belonging to $\{\phi_n,N\}$, from the formula

$$\hat{f}(\xi;\phi) = \sum_{\ell=-N}^{N} f_\ell(\xi) \exp(i\ell\phi) \qquad (4.3)$$

where the intermediate projections are in fact accurate (i.e. they are the given projections) when $\phi = \phi_n$:

$$\hat{f}(\xi;\phi_n) = f(\xi;\phi_n) \quad , \quad n \in \{1,2\ldots,N\}. \tag{4.4}$$

Note that the $f_\ell(\xi)$ are found by substituting (4.4) into (4.3). The N given projections can be transformed into what are here called *semi-projections*, of which there are 2N. Note that

$$f(-\xi;\phi) = f(\xi;\phi + \pi) \ . \tag{4.5}$$

For n > N, the semi-projections are defined by

$$f(\xi;\phi_n) = f(-\xi;\phi_n - \pi) \quad , \quad n = N+1,N+2\ldots2N \ . \tag{4.6}$$

The coefficients $f_\ell(\xi)$ are straightforwardly determined from the samples using an FFT algorithm.

Figures 4.4 and 4.5 illustrate the application of method II to the test phantom. The eight given projections have been expanded to 128 projections, 120 of which are intermediate (as defined above) and are of course approximate. The error criteria for this reconstruction are $\varepsilon(II,\Xi) = \cdot 51$ and $\varepsilon(II,\Lambda) = \cdot 85$. The large oscillatory errors which are evident towards the periphery of the image obtained by method I (Figure 4.3) are seen to be reduced in the reconstruction obtained by method II (Figure 4.5). However the images of the three small outer discs are smeared in the θ-direction (Figure 4.4) and there is only a small improvement in the central region of the image.

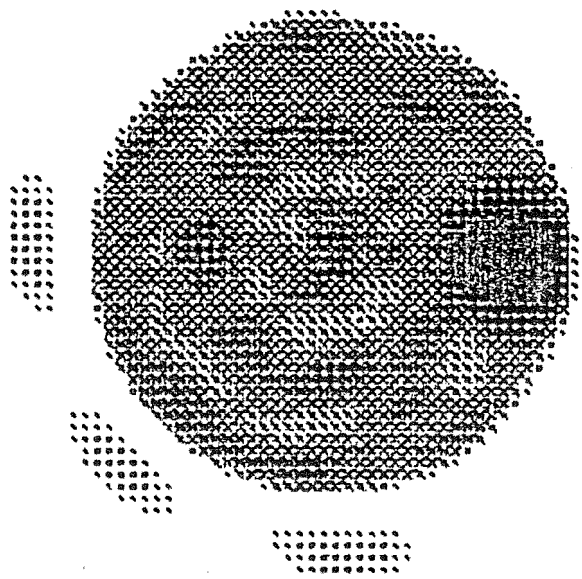


Figure 4.4 Reconstruction of the test phantom using method II (Fourier series interpolation from the eight given projections). The grey-scale covers the range of 0 (white) to 100 (black).

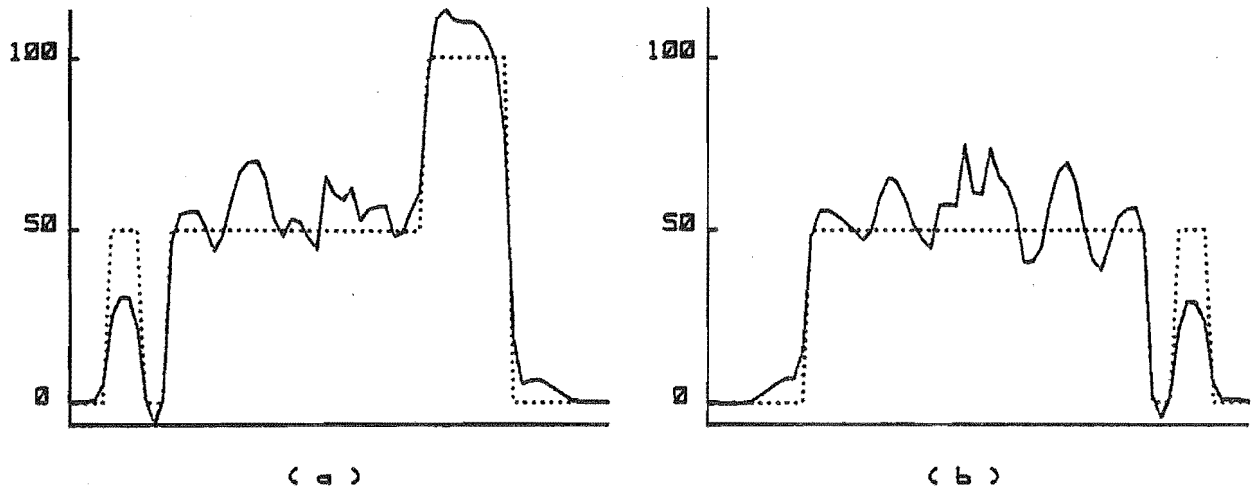


Figure 4.5 Density cuts through the test phantom (•••) and its reconstruction by method II (——): (a) cut X-X, (b) cut Y-Y.

An interpolation scheme somewhat similar to FSI is what is here called *linear interpolation*. Suppose it is desired to interpolate at M equally spaced angles between each of the given projections. For the m^{th} intermediate projection between ϕ_n and ϕ_{n+1}, the angle is defined to be $\phi_{n,m} = \phi_n + m[\phi_{n+1} - \phi_n]/(M+1)$ and the intermediate projection is itself defined to be

$$\hat{f}(\xi;\phi_{n,m}) = ((M+1-m)/(M+1))\ f(\xi;\phi_n) + (m/(M+1))\ f(\xi;\phi_{n+1}). \qquad (4.7)$$

Equation (4.7) characterises an algorithm for estimating intermediate projections. This linear interpolation scheme was applied to the test phantom and the results were virtually indistinguishable from those obtained with FSI. I decided therefore that it was unnecessary to make an independent study of linear interpolation since nothing in my computational experience suggested that significantly different results are likely to be obtained with linear interpolation compared with FSI. The following argument shows the limitations of an interpolation scheme such as linear or FSI: consider Figure 4.6 which shows the trajectory in $\xi;\phi$-space of one point in object space. The point follows a sinusoidal path. In practice the object can be considered to be composed of many points, so that projection space contains many sinusoidal paths of the form $\xi = \sin(\phi+\beta)$, where the phase β is different for each path. Any projection value $f(\xi;\phi)$ can be considered to be the sum of the contributions from all paths that cross at that value of ξ and ϕ. Of course, these trajectories cannot be isolated in $\xi;\phi$-space, but if they could be then the interpolated values could be calculated along the sinusoidal paths. Thus method II produces comparatively poor results

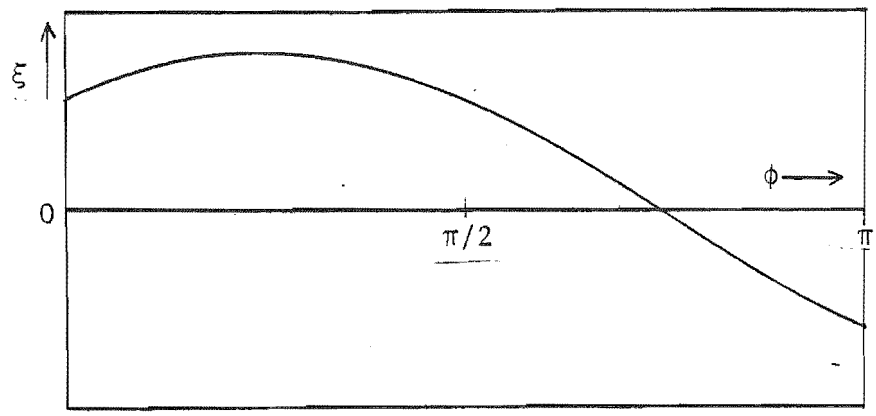


Figure 4.6 The $\xi;\phi$-plane, showing the trajectory of a point object.

because it interpolates along lines of constant ξ rather than along the sinusoidal paths, which are in general unknown *a priori*. In general it would be difficult to identify the paths, though Wagner (1979b) suggests that they might be found by using pattern recognition techniques. This only seems feasible if the object is made up of isolated dense points.

Having made the above observation, it is worth emphasising that certain useful *a priori* information concerning the image is often available. For example the reconstructed image is usually required to be real and non-negative; the size of the image - i.e. the region of image space to which the reconstructed image should be confined - can often be estimated beforehand with useful accuracy; various statistical properties of an image may also be known. The available *a priori* information is embodied in a mathematical model for the image. A state-of-the-art review of the use of mathematical models in image processing is given by Jain (1981). In the remainder of this chapter, methods are investigated which use *a priori* information about the image to aid the reconstruction.

The methods presented in §§4.3 and 4.5 are iterative. They require successive calculation of projections from an estimated image and back-projection to form a new image. Some of the numerical difficulties associated with such schemes are discussed in §4.2 and several possible solutions are presented. A reconstruction scheme based on Gerchberg's (1974) algorithm is developed in §4.3. §4.4 shows how a reconstruction algorithm can be derived by considering the problem in terms of optimisation theory. Several methods for solving optimisation problems are discussed in §4.4 and in §4.5 one of them is used to develop

a new algorithm which reconstructs an image with the aid of *a priori* information. The significance of the results is assessed in §4.6.

4.2 NUMERICAL CONSIDERATIONS

The details of the back-projection algorithm used for the examples presented in this chapter are presented in §4.2.1. Several re-projection algorithms are investigated in §§4.2.2 through 4.2.7. A tool for constraining the size of a reconstruction is introduced in §4.2.8.

4.2.1 Back-projection

In CT practice, the standard image reconstruction method is modified back-projection (see §1.6 and (1.10) and (1.15)), which involves the following sequence of operations. First, a projection $f(\xi;\phi)$ is obtained (i.e. measured, in practice) and is stored in computer memory as samples spaced by, say, Δ in the ξ-direction. The projection is then modified - by a "filtering" procedure carried out either directly (usually with dedicated hardware) or via a pair of Fourier transforms (Lewitt *et al.*, 1978). The modified projection $g(\xi;\phi)$ is now back-projected, an operation which can be visualised as follows. A rectangular grid of points, spaced by Δ in both the ξ- and η-directions, is set up in image space. The coordinates of the μ,ν^{th} point are $\xi = \mu\Delta$ and $\eta = \nu\Delta$, where μ and ν are integers. The contribution to $\lambda(r;\theta)$ at the μ,ν^{th} point, from the particular $f(\xi;\phi)$ being considered is $g(\mu\Delta;\phi)$ - i.e. each sample of the modified projection is "back-projected" along a line parallel to the η-axis. The total reconstructed density is built up from many projections, each at a different angle. Consequently, it is necessary to estimate the value of $g(\xi;\phi)$ at each point - characterised by $x = m\Delta$ and $y = n\Delta$, where m and n are integers - on a rectangular grid with sides parallel to the x- and y-axes. So, for each value of ϕ it is necessary to interpolate from the ξ,η-grid to the x,y-grid.

The modified back-projection procedure is customarily carried out using a crude interpolatory approach, such as linear-interpolation in which $g(\xi;\phi)$, at the point $x = m\Delta$, $y = n\Delta$ (this point is denoted by Q for the moment) on the x,y-grid, is approximated by a weighted average of its values (denoted for the moment by V_1 and V_2) at the two points (denoted for the moment by P_1 and P_2 respectively) on the ξ,η-grid nearest to Q. The distances QP_1 and QP_2 are written as d_1 and d_2 respectively.

The actual form of the weighted average is $[V_1 d_2 + V_2 d_1]/(d_1 + d_2)$. Consequently, linear-interpolation can be regarded as a displacement of the samples of $g(\xi;\phi)$ from their correct positions on the ξ,η-grid.

4.2.2 Re-projection

Re-projection is the process of calculating a set of projections $\hat{f}(\xi;\phi)$ from an estimate of the image $\hat{\lambda}(r;\theta)$. A re-projection scheme of high accuracy is required for the reconstruction algorithms which are developed in §§4.3 and 4.5.

The estimate of the image $\hat{\lambda}(r;\theta)$ can be visualised in two ways. The first considers that the image is constructed from small squares (*pixels*) within each of which the density is constant and which cover image space, leaving no gaps. The second considers that the image is represented by delta functions at the grid points of the x,y-grid.

A beam of width w is visualised (cf. §1.6.6 and Figure 1.7). Each sample of the projection $\hat{f}(\xi;\phi)$ contains contributions from the image $\hat{\lambda}(r;\theta)$ which lies within this beam (see Figure 4.7). The way in which each sample of $\hat{\lambda}(r;\theta)$ within this beam contributes to $\hat{f}(\xi;\phi)$ characterises a re-projection scheme.

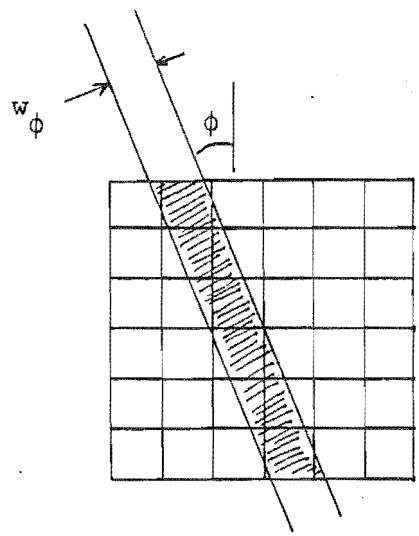


Figure 4.7 An example of a beam through a 6×6 pixel image.

The following subsections are concerned with the choice of a re-projection scheme. Many re-projection schemes have been developed over the past decade. Three schemes, designated a through c for convenience, were investigated before a satisfactory re-projection scheme was chosen. All three schemes are based on the concept that the image

can be considered as an array of numbers existing at points on an x,y-grid. The first two, which are presented in §§4.2.3 and 4.2.4, were extracted from the literature, while the third, which is presented in §4.2.5, was devised in order to investigate whether it was possible to improve on the other two. §4.2.6 shows how schemes a, b and c were tested before a satisfactory re-projection scheme was chosen. §4.2.7 presents two re-projection schemes which are due to Herman *et al.* (1973), which are discussed mainly for completeness. There is sufficient detail in §4.2.7 for it to be clear to the reader that those schemes would have functioned satisfactorily (although one of them would be computationally expensive), but in my opinion neither of them is as convenient as the scheme actually selected in §4.2.6.

4.2.3 Re-projection using linear interpolation

The first re-projection scheme (a) is the inverse of the back-projection scheme discussed in §4.2.1. Thus linear-interpolation is employed to transfer samples from the x,y-grid to the ξ,η-grid. Linear-interpolation displaces samples by $\Delta/2$ at most, which is acceptable for a back-projection routine, because Δ which is the sampling interval in the ξ-direction is almost always small enough to satisfy whatever sampling constraints apply in a particular application. But the situation changes when it is required to recompute projections from such reconstructed densities, as in the iterative schemes examined in §§4.3 and 4.5. There is the possibility of the reconstructed images being degraded by cumulative numerical errors. This consideration is illustrated by the following simple example.

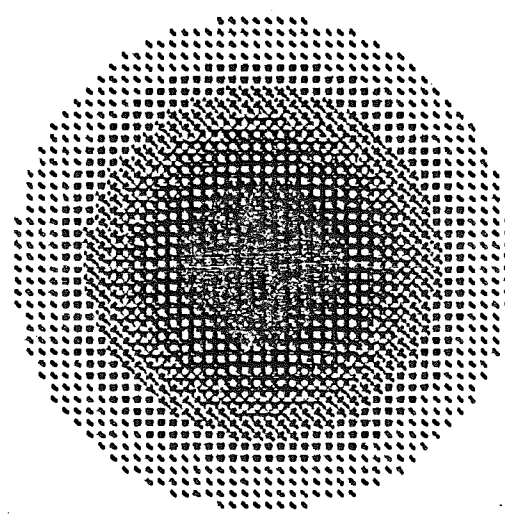


Figure 4.8 True image of the Gaussian object displayed at points on the x,y-grid. The grey-scale covers the range of 0 (white) to 100 (black).

Figure 4.8 depicts the density of a Gaussian object - if the most dense point is placed at the origin of image space then $\lambda(r;\theta)$ is proportional to $\exp(-r^2/R^2)$, where R is a real, positive constant. Note that Figure 4.8 is not a continuous tone image. Each pixel is displayed as a grey level at the appropriate point on the x,y-grid. This is done to emphasise that the errors due to the re-projection schemes are caused by the finite sampling of the image and the projections. The image exists at points on an x,y-grid which is spaced by Δ in both the x- and y-directions. Figure 4.9 shows a relief plot of projection space when re-projection scheme (a) is used to compute projections from the image which is shown in Figure 4.8. Note the large oscillatory errors at $\pi/4$ and $3\pi/4$ (cf. Peters, 1981). For this example the projection was sampled at a spacing of Δ in the ξ-direction. However the x,y-grid spacing varies with ϕ when viewed from the projection angle ϕ. The interaction of these two slightly different sampling intervals causes a moiré effect which is evident as artefacts in Figure 4.9.

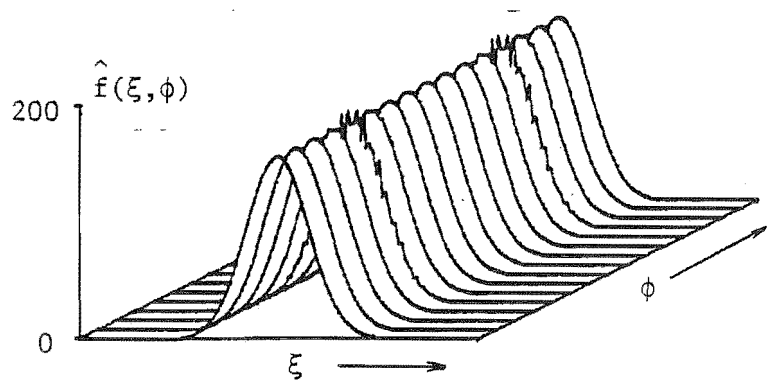


Figure 4.9 Relief plot of projection space, showing the projections calculated by scheme (a) from the image which is displayed in Figure 4.8.

Figure 4.10 is the reconstruction from the 16 projections which are shown in Figure 4.9. The streaks at 45° and 135° are large reconstruction errors which would build up if this re-projection scheme was used in an iterative reconstruction method.

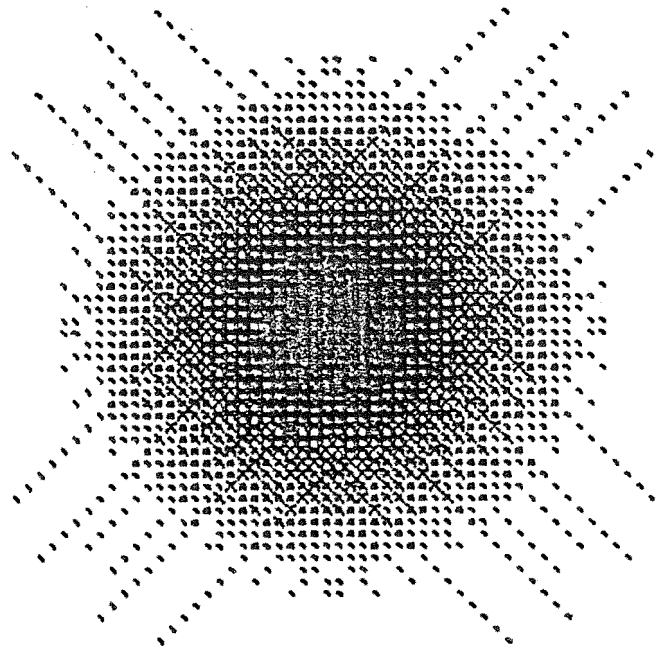


Figure 4.10 Image of the Gaussian object reconstructed from the 16 projections which are shown in Figure 4.9. The grey-scale covers the range of 0 (white) to 100 (black).

4.2.4 Re-projection due to Peters

The second re-projection scheme (b) is due to Peters (1981). It consists of scheme (a) followed by division of the projections by a *correction function* $c(\xi;\phi)$. The correction function takes account of the artefacts introduced by scheme (a). Thus $c(\xi;\phi)$ is independent of the image and so is pre-computed and stored for use wherever projections are to be calculated from an image. For convenience the (erroneous) projections calculated by scheme (a) are denoted as $\tilde{f}(\xi;\phi)$. This notation allows scheme (b) to be described concisely as:

$$\hat{f}(\xi;\phi) = \tilde{f}(\xi;\phi) \;/\; c(\xi;\phi) \;. \qquad (4.8)$$

The correction function $c(\xi;\phi)$ is calculated in the following way. Consider the smallest circle that is known to enclose the image and say its diameter is D. For each angle ϕ for which projections are to be calculated (and for which $c(\xi;\phi)$ is therefore required), a square of side D and of uniform density 1/D is set up on an x,y-grid and aligned at an angle ϕ to the grid. Since the density of the square is known, the error-free projection $\hat{f}(\xi;\phi)$ is also known (it is unity). The projection of the square is computed at this angle ϕ by scheme (a) to

give $\tilde{f}(\xi;\phi)$. This process is illustrated in Figure 4.11 for $\phi = \pi/4$. The projection $\tilde{f}(\xi;\phi)$ contains the errors introduced by scheme (a). Thus (4.8) shows that the correction function $c(\xi;\phi)$ at this particular angle ϕ is given by $\tilde{f}(\xi;\phi)/\hat{f}(\xi;\phi)$. This process is repeated for all desired angles of ϕ.

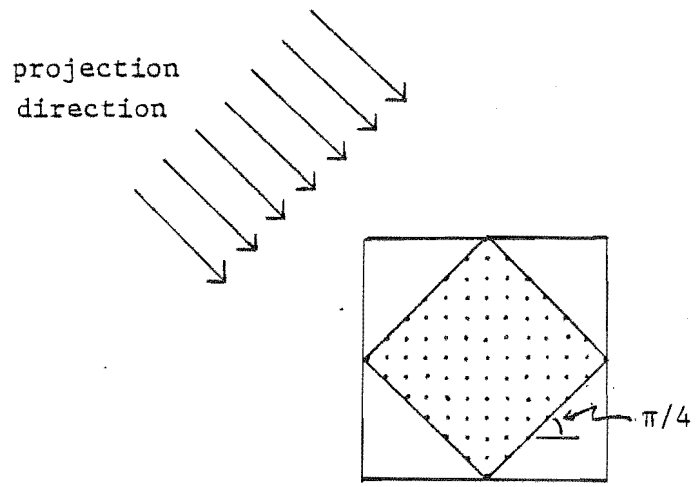


Figure 4.11 The representation on an x,y-grid of a square orientated at the projection angle ϕ (illustrated for $\phi = \pi/4$).

Figure 4.12 is a relief plot of the correction function sampled at the 16 angles required to correct the projections displayed in Figure 4.9. Scheme (b) has been applied to the example of the gaussian density distribution. The resultant projections, when modified and back-projected, produce an image which is indistinguishable from Figure 4.8.

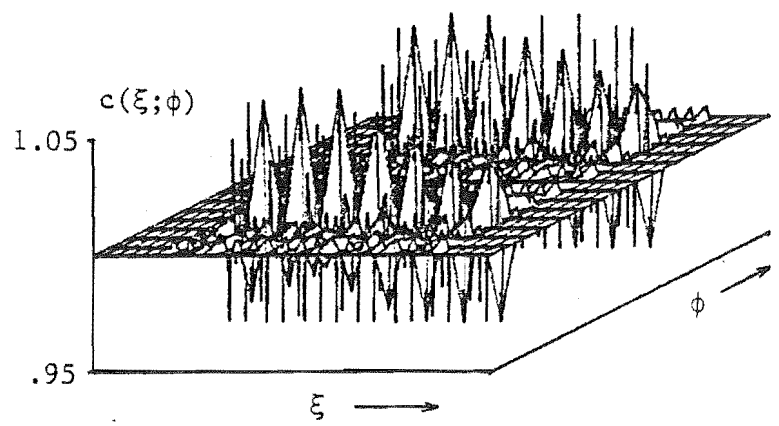


Figure 4.12 Relief plot of the correction function sampled at 16 equally spaced angles throughout π.

4.2.5 Re-projection along grid lines

For $\phi = 0^{\circ}$, 90° and 180° there is no re-projection error because the beam is parallel to one set of grid lines and perpendicular to the other. It would obviously be desirable if the beam could always be aligned along the grid lines, but of course it cannot if a single grid is used. This has suggested a scheme (c) which employs N/2 grids, from each of which is calculated two projections $\pi/2$ apart. Consider the image required to calculate projections at an angle ϕ (and $\phi + \pi/2$). To enable the projections to be calculated along the grid lines the image must be rotated at an angle of $(-\phi)$ relative to the x,y-grid. To implement this in a practical iterative scheme, N/2 images are formed at each iteration, each being the back-projection of all N modified projections, but rotated relative to the x,y-grid. Scheme (c) has also been used to calculate the projections of the gaussian density distribution shown in Figure 4.8. The resultant projections, when modified and back-projected, give an image which is indistinguishable from Figure 4.8.

4.2.6 Comparison of re-projection schemes

The three re-projection schemes described in §§4.2.3 through 4.2.5 are now compared qualitatively by incorporating each of them into an iterative scheme. Each re-projection scheme is used in turn to calculate projections from the gaussian density distribution which is illustrated in Figure 4.8. These projections are then modified and back-projected to form a new estimate of the image. The re-projection scheme is then applied to this estimate to compute a new set of projections. The process is repeated many times. The rates of propagation of errors permit the re-projection schemes to be compared. Table 4.2 lists the root-mean-square reconstruction errors for the three methods. The given projections (used to form the first image) were calculated analytically from the gaussian density distribution illustrated in Figure 4.8. Re-projection scheme (c) is seen to exhibit the lowest error. It is, however, the slowest method. For Peters' scheme (b), once the correction function is calculated at the required angles, each iteration is fast, requiring only one more multiplication per projection sample than the uncorrected scheme (a). Scheme (b) is seen to be only slightly less accurate than (c) and is invoked for the examples presented in §§4.3 and 4.5 because of the time saving for large numbers of iterations.

Table 4.2 Comparison of the r.m.s. reconstruction errors for three re-projection methods. The phantom is the gaussian density distribution. For this example there are 16 (N) projections.
Method a: uncorrected re-projection method.
Method b: Peters' correction scheme.
Method c: Projections calculated from eight (N/2) images.

Iteration	(a)	(b)	(c)
1	37.9	37.9	37.9
2	124.6	38.0	37.1
3	311.9	52.1	36.9
4			40.5

The above discussion highlights the advantage of using phantoms whose projections can be calculated analytically, when evaluating iterative schemes. The given projections are then free of interpolation error. Since the test phantom, shown in Figure 4.1, is made up of simple geometrical shapes, its projections are described by elementary mathematical formulas.

4.2.7 Two other re-projection schemes

Two re-projection schemes due to Herman *et al.* (1973) are presented in this sub-section, for completeness and to show how the re-projection problem has been attacked by what is perhaps the leading group in CT image reconstruction. Because neither of these schemes is as convenient as Peters' method, they have not been compared numerically to it.

The first scheme suggested by Herman *et al.* (1973) is based on the image being constructed of many square pixels within each of which the density is constant (cf. Figure 4.7). The pixels are each centered at the points of a square grid. The points are indexed by integers j and k. The area of intersection of a given beam and the pixel with indices j,k is denoted by a_{jk}. If the density represented by this pixel is $\hat{\lambda}_{jk}$ then its contribution to that beam is $a_{jk}\,\hat{\lambda}_{jk}$. This scheme is computationally expensive since it requires the evaluation of the area of intersection of each beam with each pixel for each projection angle.

The second scheme suggested by Herman *et al.* (1973) is based on the concept that the estimated image $\hat{\lambda}(r;\theta)$ is calculated at points on

an x,y-grid. At each angle ϕ it is desired to interpolate from the x,y-grid to the ξ,η-grid. Herman *et al.* (1973) use nearest-neighbour interpolation in which $\hat{\lambda}(r;\theta)$ at the point $\xi = \mu\Delta$, $\eta = \nu\Delta$ (this point is denoted by Q for the moment) on the ξ,η-grid, is approximated by the sample on the x,y-grid nearest to Q. The sample of the projection at $\xi = \mu\Delta$ is found by summing along the grid line of constant ξ. This scheme is similar to that described in §4.2.3 except that it uses nearest-neighbour rather than linear interpolation when transferring samples from the x,y-grid to the ξ,η-grid. Thus in common with scheme (a) this scheme gives oscillatory errors on the re-computed projections. As discussed in §4.2.3 these errors are a moiré artefact which occurs because the spacing of the x,y-grid varies relative to the ξ-spacing (the beam width) as the ξ,η-grid rotates. To get round this Herman *et al.* (1973) suggest varying the width of the beam w_ϕ with projection angle. Thus the spacing of the ξ,η-grid is dependent on ϕ. They let $w_\phi = w|\cos\phi|$ for $0 \leq |\phi| \leq \pi/4$ and $w_\phi = w|\sin\phi|$ for $\pi/4 < |\phi| \leq \pi/2$ where w is a constant. They have used this re-projection scheme with the ART algorithm and report that letting w_ϕ vary with ϕ gives superior results to holding it constant.

4.2.8 Extent of the image

The algorithms developed in §§4.3 and 4.5 are predicated on the image being real and non-negative (this is the condition defined as "image positivity" in §1.3). Another constraint which might usefully be incorporated in such schemes is the *extent* of the image, which is defined to be the smallest convex region (or regions) in image space which encloses the image. This concept of extent means that if the object consists of several regions which are not connected, then the extent is constructed from separated convex regions. An estimate of the extent of the image is often available in practice. This sub-section is concerned with one way of obtaining an estimate of the extent.

A *null ray* is defined to be any straight line through image space along which the integrated density is zero, i.e. for the particular values of ξ and ϕ characterising that line, it follows that $f(\xi;\phi) = 0$. Since the image density cannot be negative, the image must be zero along the null ray (this is pointed out by Brooks and Di Chiro, 1976, §4.6). It is therefore useful to define an image *mask* as follows. Generate an x,y-grid which is pre-set to some arbitrary (non-zero) level. For each

null ray, set to zero the points on the x,y-grid which contribute to it. Repeat this for all null rays. Figure 4.13 shows the image mask obtained from eight given projections of the test phantom. Wherever the mask is zero (white in Figure 4.13), the density of the reconstructed image should also be zero. Thus the image mask gives an estimate of the maximum extent of the image. Of course any "dents" in the image cannot be found by this method.

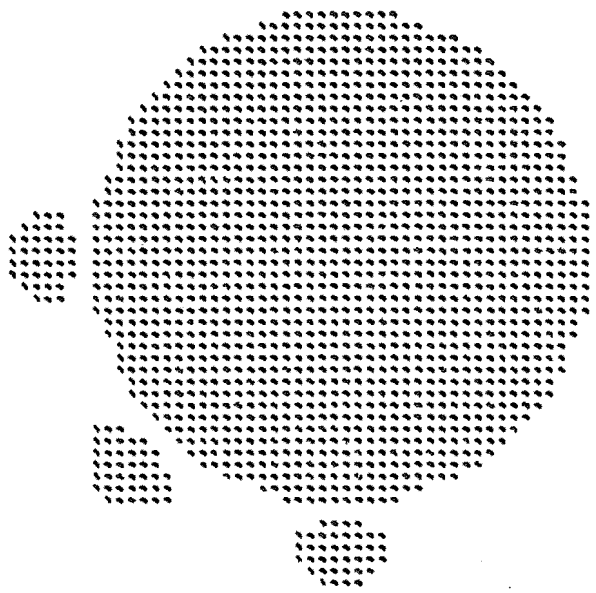


Figure 4.13 Image mask obtained from eight given projections of the test phantom.

4.3 GERCHBERG'S ALGORITHM

In computed tomography, measurements are made in image space. However, the projection theorem (1.4) shows that the one-dimensional Fourier transform of each projection lies along a radial line in Fourier space. So, it is valid to consider that the available measurements are of the spectrum (spatial frequency) of the image. When few projections are given, the spectrum is incomplete. The algorithm due to Gerchberg (1974), which is discussed in this section, is concerned with extrapolating from the given part of the spectrum into the region of Fourier space for which there are no available measurements.

It is convenient to denote the measured part of the spectrum as F_{meas} and that part of the spectrum for which an estimate is to be generated as F_{ext}. Thus the complete spectrum $F = F_{meas} + F_{ext}$.

Gerchberg's algorithm is an iterative scheme which successively applies constraints in the frequency domain and in image space. In order to start the scheme an initial guess must be made for F_{ext} (it is often chosen to be zero). The first step is to take the Fourier transform of the spectrum F which provides an estimate $\hat{\lambda}$ of the image. At this stage certain *a priori* knowledge is used to correct $\hat{\lambda}$. It is known that the image density cannot be negative, so any parts of $\hat{\lambda}$ which are negative are set to zero. An estimate of the extent of the image may also be available (cf. §4.2.8). Thus $\hat{\lambda}$ is also set to zero outside the estimated extent. This gives a corrected estimate $\tilde{\lambda}$ of the image. The next step is to take the inverse Fourier transform which gives an estimate $\hat{F}$ of the complete spectrum of the image. Now $\hat{F}$ can be written as $(\hat{F}_{meas} + \hat{F}_{ext})$, i.e. the scheme has provided a new estimate for both F_{meas} and F_{ext}, where $\hat{F}_{meas}$ is the new estimate for the measured part of the spectrum and $\hat{F}_{ext}$ is the new estimate for the extrapolated region of the spectrum. The final step is to replace the estimate $\hat{F}_{meas}$ with the measured part of the spectrum F_{meas}. This gives a new estimate for the complete spectrum $F = F_{meas} + \hat{F}_{ext}$. This iteration scheme is repeated until a *stopping criterion* is satisfied. A typical stopping criterion could be based on image positivity - for example the iterations would cease when the minimum density in $\hat{\lambda}$ rises above some predetermined negative level.

Sato *et al.* (1981) and Tam and Perez-Mendez (1981) use Gerchberg's algorithm to reconstruct images when the given projections do not uniformly cover the range of $\phi = 0$ to π, i.e. they consider the problem caused by a sector of missing projection angles (cf. §1.6.3). This section presents two methods based on Gerchberg's algorithm which are designed to reconstruct an image when there is only a small number of equally-spaced projections. It is desired to estimate extra projections so that there is a complete set of M projections ($M > N$). The algorithm estimates $\{\hat{f},\psi_m,M\}$ where $\psi_m = m\pi/M$ and

$$\hat{f}(\xi;\psi_m) = f(\xi;\phi_n) \quad \text{when} \quad \psi_m \in \{\phi_n,N\} . \tag{4.9}$$

A full set of M projections is required to start the iteration scheme. Two choices are illustrated here for the initial guess at the projections $\hat{f}(\xi;\psi_m)$ where $\psi_m \notin \{\phi_n,N\}$. The first assumes that they are zero, i.e.

$$\hat{f}(\xi;\psi_m) = 0, \quad \psi_m \notin \{\phi_n,N\} . \tag{4.10}$$

The second uses the projections estimated by FSI (4.3) as the initial guess. A reconstruction scheme based on Gerchberg's algorithm is now described:

(i) Modify and back-project the full set of projections $\{\hat{f},\psi_m,M\}$ to give $\hat{\lambda}(r;\theta)$.

(ii) Cease the iterations when $\hat{\lambda}(r;\theta)$ stops converging towards whatever *a priori* constraint is being applied.

(iii) Form $\tilde{\lambda}(r;\theta)$ by introducing the *a priori* information into $\hat{\lambda}(r;\theta)$.

(iv) Calculate a full set of projections $\{\hat{f},\psi_m,M\}$ from $\tilde{\lambda}(r;\theta)$.

(v) Replace the given projections $\{f,\phi_n,N\}$ within this set.

(vi) Continue until the iterations cease at step (ii).

The two methods (III and IV) based on this scheme are effectively ways of ensuring positivity. For the first method (III), those samples of $\hat{\lambda}(r;\theta)$ which are negative, are set to zero at step (iii). In the second method (IV), the image mask described in §4.2.8 is invoked. Where the mask is zero, $\hat{\lambda}(r;\theta)$ is always set to zero. The mask for eight projections of the test phantom is shown in Figure 4.13.

It was originally intended to base the stopping criterion (step ii) on image positivity. Thus a measure D of the discrepancy between the current estimate of the image $\hat{\lambda}(r;\theta)^{(\nu)}$ and the *a priori* knowledge is introduced:

$$D^{(\nu)} = - \sum_{j,k} \min\left(0, \hat{\lambda}_{jk}^{(\nu)}\right) \tag{4.11}$$

where ν is the iteration number. $D^{(\nu)}$ is the sum of the samples of the current estimate of the density which are negative. The iteration scheme could be stopped when $D^{(\nu)}$ falls below an acceptable level or when the improvement to the image at each iteration

$$D^{(\nu-1)} - D^{(\nu)} \tag{4.12}$$

becomes small enough. However for methods III and IV it was found that the discrepancy D first diminished and then increased with more iterations. Computational experience shows that the image estimate first converges towards the true image and then diverges. Thus the "best" image given by these methods occurs at iteration $(\nu - 1)$ where

$$D^{(\nu)} > D^{(\nu-1)} \quad . \tag{4.13}$$

This was used as the stopping criterion for the examples presented in this section.

Tables 4.3 and 4.4 list the error criteria for the two methods for different M when there are eight (N) given projections. The results recorded in Table 4.4 are superior to those recorded in Table 4.3 in all cases. Both methods are based on the same *a priori* information, yet method IV gives better results. For both methods, the lowest error within Λ (defined in §4.1) is given when (4.10) defines the projections for step (i).

Table 4.3 Comparison criteria for method III applied to the test phantom. The eight given projections are expanded to M projections. The initial guess for the unknown projections is either that they are zero (0) or that they are those given by method II (FS).

M	Initial guess	Number of iterations	$\varepsilon(III,\Xi)$	$\varepsilon(III,\Lambda)$
16	0	1	.64	.76
16	FS	2	.61	.84
32	0	1	.56	.76
32	FS	7	.51	.80
64	0	2	.56	.79
64	FS	4	.50	.84

Table 4.4 Comparison criteria for method IV applied to the test phantom. The eight given projections are expanded to M projections. The initial guess for the unknown projections is either that they are zero (0) or that they are those given by method II (FS).

M	Initial guess	Number of iterations	$\varepsilon(IV,\Xi)$	$\varepsilon(IV,\Lambda)$
16	0	1	.50	.71
16	FS	3	.53	.76
32	0	3	.41	.71
32	FS	11	.43	.73
64	0	3	.42	.74
64	FS	4	.46	.80

Tables 4.3 and 4.4 also show that convergence is slower when the initial guess is given by (4.3). These observations are somewhat surprising since the projections given by (4.3) are closer to the "correct" projections than those given by (4.10). This might be explained by the fact that (4.10) does not bias the interpolation scheme towards a particular solution .

For both methods, $M = 32$ gives better results than $M = 64$. The reason for this could be that image positivity does not provide sufficient added information. In fact, Table 4.4 shows that when (4.10) is employed, the error within Λ for $M = 16$ is the same as for $M = 32$. The error outside Λ is lower, however, when $M = 32$.

Figures 4.14 and 4.15 illustrate the best result obtained by methods III and IV. The grey-scale picture of Figure 4.14 is the reconstruction obtained by method IV, for $M = 32$ and when (4.10) is employed. Density cuts through this image are presented in Figure 4.15. Comparison of Figure 4.15 with Figure 4.5, and Figure 4.14 with Figure 4.4 shows that method IV (Gerchberg's algorithm) produces reconstructions which are superior to Fourier series interpolation (method II).

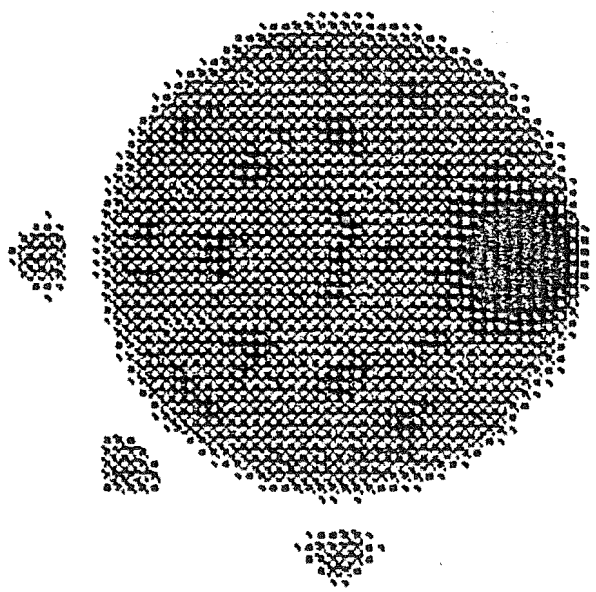


Figure 4.14 Reconstruction of the test phantom given by method IV (Gerchberg's algorithm) after three iterations with $N = 8$, $M = 32$ and using (4.10) as the initial guess. The grey-scale covers the range of 0 (white) to 100 (black).

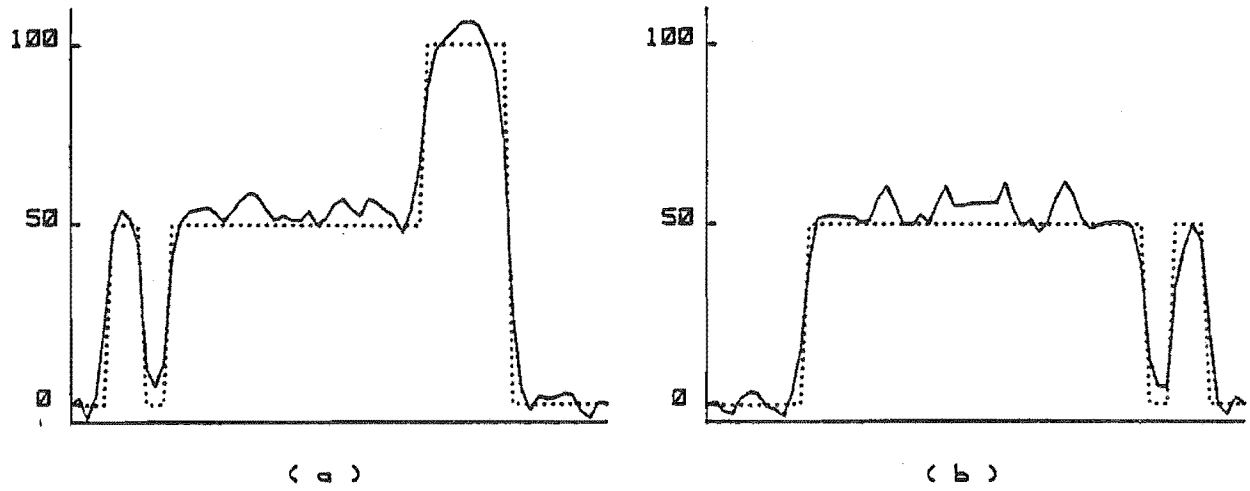


Figure 4.15 Density cuts through the test phantom (•••) and its reconstruction by method IV (——): (a) cut X - X, (b) cut Y - Y.

4.4 OPTIMISATION TECHNIQUES FOR IMAGE RECONSTRUCTION FROM PROJECTIONS

A general optimisation problem can be posed as: find that $\underline{x}$ which maximises $F = F(\underline{x})$, where $\underline{x}^T = [x_1\ x_2 \ldots x_n]$ and $\underline{x}$ is a column vector of n system parameters. The aim of optimisation is to determine the parameters $\underline{x}_{max}$ which give the maximum value E_{max} of the *objective function* F. It is initially considered here that there are no constraints on $\underline{x}$ - i.e. it is an *unconstrained optimisation* problem. A point $\underline{x}_{max}$ which gives the highest possible value of F is termed a *global maximum*. In general it need not be unique. In practice it is very difficult to determine if the maximum obtained by a numerical process is a global maximum or not. In most circumstances it can only be said that the maximum obtained is a maximum within a local area of search. The point $\underline{x}_{max}$ is therefore termed a *local maximum*.

The simplest class of methods used to solve optimisation problems are known as *search methods*. These require only evaluations of $F(\underline{x})$. The values of $F(\underline{x})$ for several test values $\underline{x}$ are used to steer the search for $\underline{x}_{max}$. Algorithms such as the Fibonacci Search and the Golden Section Search can be used to determine the positions of the test points for the search (cf. Adby and Dempster, 1974). If the function F has continuous derivatives it is found that methods which utilise the derivatives converge to the optimal $\underline{x}_{max}$ quicker than methods which employ evaluations of only the function itself.

Gradient methods for optimisation are based on the *Taylor expansion*:

$$F(\underline{x} + \Delta\underline{x}) = F(\underline{x}) + g^T(\underline{x})\,\Delta\underline{x} + \tfrac{1}{2}\,\Delta\underline{x}^T H(\underline{x})\,\Delta\underline{x} + \ldots \tag{4.14}$$

where

$$\Delta\underline{x}^T = [\Delta x_1\ \Delta x_2\ \ldots\ \Delta x_n], \tag{4.15}$$

and

$$g^T(\underline{x}) = \nabla F(\underline{x}) = \left[\frac{\partial F(\underline{x})}{\partial x_1}\ \frac{\partial F(\underline{x})}{\partial x_2}\ \ldots\ \frac{\partial F(\underline{x})}{\partial x_n}\right] \tag{4.16}$$

which is called the *Jacobian gradient vector*. Note that the symbol g is also used for the modified projections. However when there is likely to be confusion, the text clarifies the distinction. The n×n Hessian matrix is defined by

$$\{H(\underline{x})\}_{jk} = \frac{\partial^2 F(\underline{x})}{\partial x_j\,\partial x_k}\ . \tag{4.17}$$

The maximum of the function $F(\underline{x})$ occurs at a stationary point. Therefore, the optimisation problem can be restated as: find the root $\underline{x}_{max}$ of

$$g(\underline{x}_{max}) = \underline{0}\ . \tag{4.18}$$

This problem can be solved as a *Picard iteration scheme* (cf. Isaacson and Keller, 1966):

$$\underline{x}^{(\nu+1)} = \underline{x}^{(\nu)} - A(\underline{x})\ g(\underline{x}^{(\nu)}) \tag{4.19}$$

where $A(\underline{x})$ is an n^{th} order square matrix with components $a_{jk}(\underline{x})$, and ν is the iteration number. The simplest choice for $A(\underline{x})$ is

$$A(\underline{x}) = A\ , \tag{4.20}$$

where A is a constant non-singular matrix. The aim of such a scheme is to move successively closer to the optimum $\underline{x}_{max}$. $\underline{x}^{(\nu)}$ is the *current estimate* of the optimum and $\underline{x}^{(\nu+1)}$ is the next estimate. Thus the term $A(\underline{x})\ g(\underline{x}^{(\nu)})$ is the *step* which gives the next estimate from the current one. The algorithm steps towards the optimum state and so such algorithms are often described as "hill-climbing".

The common *steepest ascent* algorithm is an example of a hill climbing solution to the optimisation problem. Consider the first order approximation given by the Taylor expansion (4.14) to the effect of a change $\Delta\underline{x}$ in $\underline{x}$:

$$\Delta E = g^T(\underline{x})\ \Delta(\underline{x}) = \sum_{k=1}^{n} \frac{\partial F(\underline{x})}{\partial x_k}\ \Delta x_k\ . \tag{4.21}$$

Equation (4.21) is a maximum when $\Delta\underline{x}$ is in the direction of the gradient vector $g(\underline{x})$. The unit vector in the direction of $g(\underline{x})$ is given by $\underline{u}$ where

$$\underline{u} = g(\underline{x}) \; / \; |g(\underline{x})| \tag{4.22}$$

and the (*Euclidian*) *magnitude* or *norm* $|g(\underline{x})|$ is a number given by the square root of the sum of the squares of the elements of $g(\underline{x})$, i.e.

$$|g(\underline{x})| = \left(\sum_{k=1}^{n} (g_k(\underline{x}))^2 \right)^{\frac{1}{2}} = \left(\sum_{k=1}^{n} \left(\frac{\partial F(\underline{x})}{\partial x_k}\right)^2 \right)^{\frac{1}{2}} . \tag{4.23}$$

Thus the hill climbing algorithm can be described by

$$\underline{x}^{(\nu+1)} = \underline{x}^{(\nu)} + \beta g(\underline{x}^{(\nu)}) \tag{4.24}$$

where β is the *step length*. Usually β is determined by a single parameter or linear search to maximise the function of β defined by $F(\underline{x}^{(\nu)} + \beta\, g(\underline{x}^{(\nu)}))$. An alternative is to consider steps that are a constant fraction of the gradient vector. This simplifies programming and reduces the number of evaluations of $F(\underline{x})$ and $g(\underline{x})$ at each iteration. However, the number of iterations is generally increased and convergence cannot be guaranteed unless the fraction is chosen carefully.

There are *improved gradient methods* which rely on the Hessian matrix defined by (4.17). Consider the effect of a change $\Delta\underline{x}$ which can be approximated by the terms up to second order in a Taylor expansion (4.14):

$$\Delta E = g^T(\underline{x})\, \Delta\underline{x} + \tfrac{1}{2}\, \Delta\underline{x}^T\, H(\underline{x})\, \Delta\underline{x} \tag{4.25}$$

which can be re-written as

$$F(\underline{x}_{max}) \simeq F(\underline{x}) + \sum_{k=1}^{n} \frac{\partial F(\underline{x})}{\partial x_k} \Delta x_k + \tfrac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{n} \Delta x_k \frac{\partial^2 F(\underline{x})}{\partial x_k \, \partial x_j} \Delta x_k . \tag{4.26}$$

To evaluate $\Delta\underline{x}$ approximately, $g(\underline{x})$ and $H(\underline{x})$ are considered fixed. Equation (4.26) is differentiated with respect to the elements Δx_j for each j from 1 to n and the results are set to zero:

$$\frac{\partial F(\underline{x})}{\partial x_j} + \sum_{k=1}^{n} \Delta x_k \frac{\partial^2 F(\underline{x})}{\partial x_k \, \partial x_j} = 0 \quad , \quad j = 1, \ldots n . \tag{4.27}$$

In matrix form, (4.27) is re-expressed as

$$g(\underline{x}) + H(\underline{x})\, \Delta\underline{x} = \underline{0} . \tag{4.28}$$

Thus (4.28) implies that

$$\Delta \underline{x} \;=\; - H^{-1}(\underline{x})\; g(\underline{x}) \tag{4.29}$$

is the approximation to the required step to the maximum $\underline{x}_{max}$ from a point $\underline{x}$ near to the maximum (the current point). Equation (4.29) is fundamental to all second order solutions to the optimisation problem. When it is used directly to generate successive steps towards a maximum from an initial state $\underline{x}^{(o)}$ the method is known as the generalised *Newton-Raphson* method (or simply as *Newton's* method). Direct use of (4.29) is limited however, because the Hessian matrix H must be computed and inverted at each iteration. This is generally computationally expensive and since the storage requirements increase with n^2 such methods are impractical for large problems.

The Hessian matrix can be used to develop a first order scheme which does not require a search for the step length β. The increment $\Delta \underline{x}$ (= $\beta\, g(\underline{x})$ from (4.24)) is first substituted into (4.25) so that

$$\Delta E \;=\; \beta\; g^T(\underline{x})\; g(\underline{x}) + \tfrac{1}{2}\, \beta^2\; g^T(\underline{x})\; H(\underline{x})\; g(\underline{x}). \tag{4.30}$$

It is desired to calculate the optimal step length β which maximises (4.30). Thus

$$\frac{\partial\, \Delta E}{\partial\, \beta} \;=\; |g(\underline{x})| + \beta\; g^T(\underline{x})\; H(\underline{x})\; g(\underline{x}) \;=\; 0\;. \tag{4.31}$$

Therefore

$$\beta \;=\; \frac{|g(\underline{x})|^2}{g^T(\underline{x})\; H(\underline{x})\; g(\underline{x})}\;. \tag{4.32}$$

Combining (4.30) and (4.14) gives

$$F(\underline{x} + \beta\; g(\underline{x})) \;=\; F(\underline{x}) + \beta\; g^T(\underline{x})\; g(\underline{x}) + \frac{\beta^2}{2}\; g^T(\underline{x})\; H(\underline{x})\; g(\underline{x}) \tag{4.33}$$

Thus

$$g^T(\underline{x})\; H(\underline{x})\; g(\underline{x}) \;=\; \frac{2}{\beta^2}\; \{F(\underline{x} + \beta\; g(\underline{x})) - F(\underline{x}) - \beta\; |g(\underline{x})|^2\}\;. \tag{4.34}$$

If an initial guess $\beta^{(o)}$ is supplied, (4.32) and (4.34) can be used to update the step length:

$$\beta^{(\nu+1)} \;=\; \frac{|g(\underline{x})|^2\; (\beta^{(\nu)})^2}{2\; \{F(\underline{x} + \beta^{(\nu)}\; g(\underline{x})) - F(\underline{x}) - \beta^{(\nu)}\; |g(\underline{x})|^2\}}\;. \tag{4.35}$$

The iteration scheme specified by (4.24) and (4.35) no longer requires a search for the step length. However the method requires an initial guess $\beta^{(o)}$ and the evaluation of the function at each iteration.

In *Aitken's* δ^2 *- Method*, (4.19) is replaced by:

$$\underline{x}^{(\nu+1)} = \underline{x}^{(\nu)} - \frac{\underline{e}_{(\nu)}^{2}}{\hat{\underline{e}}_{(\nu+1)} - \underline{e}_{(\nu)}} , \tag{4.36}$$

where

$$\underline{e}_{(\nu)} = \hat{\underline{x}}^{(\nu+1)} - \underline{x}^{(\nu)} \tag{4.37}$$

is the error between the two approximations to the solution of (4.18) and

$$\hat{\underline{e}}_{(\nu+1)} = \hat{\underline{x}}^{(\nu+2)} - \hat{\underline{x}}^{(\nu+1)} \tag{4.38}$$

where $\hat{\underline{x}}^{(\nu+2)}$ is the estimate of the optimum parameters given by using $\hat{\underline{x}}^{(\nu+1)}$. The estimates $\hat{\underline{x}}$ to the optimum state are given by one of the earlier methods such as steepest ascent. Descriptions of other advanced techniques (e.g. *conjugate gradients*) can be found in standard texts on optimisation theory (cf. Adby and Dempster, 1974).

The convergence rate of the basic scheme (4.24) can be improved by modifying the Picard iteration scheme as follows (cf. Isaacson and Keller, 1966):

$$\underline{x}^{(\nu+1)} = \underline{x}^{(\nu)} + \Theta\, g(\underline{x}^{(\nu)}) \tag{4.39}$$

where Θ is a diagonal matrix of convergence factors defined by

$$\Theta = \theta_j \delta_{jk} , \quad \det [\Theta] = \theta_1 \theta_2 \ldots \theta_n \neq 0 . \tag{4.40}$$

The diagonal matrix of convergence factors is obtained from a consideration of the particular problem (cf. Trussell and Hunt, 1979).

The optimisation schemes discussed above attempt to solve the unconstrained optimisation of $F(\underline{x})$. However in most practical situations there are constraints on the allowed solutions. Consider equality constraints of the form

$$h_j(\underline{x}) = 0 , \quad j = 1,2\ldots m . \tag{4.41}$$

There are several techniques for approaching such a constrained problem directly (cf. Adby and Dempster, 1974), though the most commonly used technique is to convert the problem into one of unconstrained optimisation.

This can then be solved by one of the techniques discussed above. The classical method of removing constraints of the form of (4.41) is due to Lagrange. A new function is introduced:

$$J = F(\underline{x}) - \sum_{j=1}^{m} \gamma_j \, h_j(\underline{x}) \qquad (4.42)$$

involving the 'undetermined' *Lagrange multipliers* γ_j, $j = 1, \ldots m$. Instead of a constrained problem it is now required to find the unconstrained maximum of the right-hand side of (4.42). There are now $n+m$ variables $x_1, \ldots x_n$; $\gamma_1, \ldots \gamma_m$ to be solved for.

Any image restoration algorithm can be described as an optimisation problem (Andrews and Hunt, 1977, §6.5). Reconstructing an image from a set of measured projections can also be posed as an optimisation problem. The first step in such an approach is to define some function of the image which is to be maximised. This function is usually of the form

$$F(\lambda(r;\theta)) = \iint_{\Omega} \Phi\{\lambda(r;\theta)\} \, d\Omega \qquad (4.43)$$

where Ω is the region of image space throughout which the density is to be reconstructed and $d\Omega$ is the element of area in image space. Durrani and Goutis (1980) list seven alternatives for the functional Φ (see Table 4.5).

Table 4.5 Cost criteria for image reconstruction (Durrani and Goutis, 1980). a, $\{a_m\}$, $\{b_m\}$ are weighting constants.

	$\Phi\{A\}$	
1	$\frac{1}{2} A^2$	Least squares
2	$\ln A$	Entropy 1
3	$- A \ln A$	Entropy 2
4	A^{μ} for $\mu > 1$	
5	$\frac{1}{2}\left\{\left(\frac{\partial A}{\partial x}\right)^2 + \left(\frac{\partial A}{\partial y}\right)^2\right\}$	Gradient norm
6	$\frac{1}{2}\left[(1-a)\, A^2 + a\left\{\left(\frac{\partial A}{\partial x}\right)^2 + \left(\frac{\partial A}{\partial y}\right)^2\right\}\right]$	Least squares with a spatial filter
7	$\frac{1}{2} A^2 + \frac{1}{2} \sum_{m=1}^{n}\left[a_m\left(\frac{\partial^m A}{\partial x^m}\right)^2 + b_m\left(\frac{\partial^m A}{\partial y^m}\right)^2\right]$	Least squares with high-order spatial filter

The reconstruction of an image from a set of projections can thus be formulated as: calculate the image $\lambda(r;\theta)$ which maximises the integral in (4.43), subject to the measured projections - $\{f,\phi_n,N\}$. The constraints can be expressed as

$$\int_{-\infty}^{\infty} \hat{\lambda}(r;\theta)\, d\eta \;=\; f(\xi;\phi) \qquad (4.44)$$

where $\hat{\lambda}(r;\theta)$ is the current estimate of the image. A difference projection can be defined by re-writing (4.44) as

$$f_d(\xi;\phi) \;=\; f(\xi;\phi) - \hat{f}(\xi;\phi) \qquad (4.45)$$

where $\hat{f}(\xi;\phi)$ is given by the left-hand side of (4.44), which is the projection calculated from the current estimate of the image. In some situations it is desirable to explicitly consider the noise in the measurement process. The measurement errors are usually assumed to be independent random variables of zero mean and with variance $\sigma^2(\xi;\phi)$. There are N measured projections with M samples in each so that (4.44) could be modified to:

$$\sum_{m=1}^{M} \sum_{n=1}^{N} \left(f_d(\xi_m;\phi_n)\right)^2 \,/\, \sigma^2(\xi_m;\phi_n) \;=\; M\,N\,. \qquad (4.46)$$

There are two approaches commonly used to convert the above constrained optimisation problem into an unconstrained optimisation problem. The first of these employs one-dimensional Lagrange-multiplier functions $\gamma(\xi;\phi)$ (cf. Frieden, 1972; Minerbo, 1979; Durrani and Goutis, 1980). It enables the unconstrained optimisation problem to be expressed as: maximise

$$J \;=\; \iint_{\Omega} \Phi\{\lambda(r;\theta)\}\, d\Omega - \iint_{T} \gamma(\xi;\phi)\; f_d(\xi;\phi)\; \frac{1}{\xi}\, d\Omega \qquad (4.47)$$

where T is the region throughout which the projection data are available. The reconstruction (in principle) entails the maximisation of (4.47) to obtain the Lagrange-multiplier functions and the image density. In practice, Durrani and Goutis (1980) show that for particular choices of the functional $\Phi\{\cdot\}$, the reconstruction can be obtained without the use of an iterative scheme.

The second approach uses only one Lagrange multiplier γ, but includes the variance of the measurement errors $\sigma^2(\xi;\phi)$. The proponents of this method (Wernecke and D'Addario, 1977; Gull and Daniell, 1978) formulate the problem as the unconstrained maximisation of

$$J = \iint_{\Omega} \Phi\{\lambda(r;\theta)\}\, d\Omega - \gamma \iint_{T} \frac{[f_d(\xi;\phi)]^2}{\sigma^2(\xi;\phi)} \frac{1}{\xi}\, d\Omega \,. \qquad (4.48)$$

In this case the noise variance $\sigma^2(\xi;\phi)$ is required *a priori* for the reconstruction. The other major difference between (4.47) and (4.48) is that the latter includes $f_d(\xi;\phi)$ as a squared term. This scheme tends to minimise the "mean-square-error" rather than a "mean error" as in (4.47).

An image can be reconstructed by maximising either (4.47) or (4.48) by one of the schemes discussed earlier in this section. Practical considerations usually mean that a first order method has to be used (cf. Wernecke and D'Addario, 1977). Consider the problem of reconstructing an image of 64×64 pixels (current CT machines use a 512×512 image grid so this is a small image). Thus the aim is to caclulate n = 64×64 = 4096 variables by the optimisation scheme. The gradient vector, (4.16), is also of order n but the Hessian, (4.17), has n^2 elements. For the above example this would require calculation and storage of up to 16 777 216 numbers. Clearly, methods which employ the Hessian matrix, or approximations to it, are of no use for problems of this size. Consequently, the reconstruction method developed in the next section uses a first order optimisation technique.

4.5 A RECONSTRUCTION TECHNIQUE EMPLOYING *A PRIORI* INFORMATION

The methods developed in this section are based on the optimisation problem characterised by (4.48). Here, however, no explicit account is taken of the noise, i.e. the noise variances are assumed to be constant and are included in the multiplier γ. In order to reduce complexity, the parameter γ is considered to be a constant (predetermined) value. This approach is also followed by Baba *et al.* (1981). Thus γ is not strictly a Lagrange multiplier, but can be considered to determine the trade-off between accuracy (the second term in equation 4.48) and the *a priori* knowledge (the first term).

Optimisation of (4.48) is carried out here by the hill-climbing algorithm specified by (4.24). The elements of the gradient vector are given by:

$$\frac{\partial J}{\partial \hat{\lambda}(r;\theta)} = \frac{\partial \Phi\{\hat{\lambda}(r;\theta)\}}{\partial \hat{\lambda}(r;\theta)} - 2\gamma \int_0^{\infty} \int_0^{2\pi} f_d(\xi;\phi) \frac{\partial \hat{f}(\xi;\phi)}{\partial \hat{\lambda}(r;\theta)} \, d\phi \, d\xi . \tag{4.49}$$

The first term on the right-hand side can be calculated if the functional $\Phi\{\cdot\}$ is known analytically. The second term requires the further evaluation of a derivative. A projection of the current estimate of the image can be re-expressed as (cf. (1.3)):

$$\hat{f}(\xi;\phi) = \int_0^{\infty}\int_0^{2\pi} \hat{\lambda}(r;\theta) \, \delta(\xi - r\cos(\theta-\phi)) \, r \, d\theta \, dr \tag{4.50}$$

where $\delta(\cdot)$ denotes the Dirac delta function. Therefore

$$\frac{\partial \hat{f}(\xi;\phi)}{\partial \hat{\lambda}(r;\theta)} = \delta(\xi - r\cos(\theta-\phi)) . \tag{4.51}$$

Substituting (4.51) into (4.49) gives

$$\frac{\partial J}{\partial \hat{\lambda}(r;\theta)} = \frac{\partial \Phi\{\hat{\lambda}(r;\theta)\}}{\partial \hat{\lambda}(r;\theta)} - 2\gamma \int_0^{2\pi} f_d(r\cos(\theta-\phi);\phi) \, d\phi . \tag{4.52}$$

It is convenient to define the *difference density*

$$\lambda_d(r;\theta) = \int_0^{2\pi} f_d(r\cos(\theta-\phi);\phi) \, d\phi \tag{4.53}$$

which is the back-projection of all (N) difference projections. It should be noted that the difference projections, $f_d(\xi;\phi)$, are not modified before back-projection. Thus $\lambda_d(r;\theta)$ is not a true density - it is a *layergram* of the difference projections (cf. Bates and Peters, 1971). Using the difference density, the elements of the gradient vector can be expressed simply as

$$\frac{\partial J}{\partial \hat{\lambda}(r;\theta)} = \frac{\partial \Phi\{\hat{\lambda}(r;\theta)\}}{\partial \hat{\lambda}(r;\theta)} - 2\gamma \, \lambda_d(r;\theta) . \tag{4.54}$$

The iterative solution (4.24) can be re-expressed as

$$\hat{\lambda}(r;\theta)^{(\nu+1)} = \hat{\lambda}(r;\theta)^{(\nu)} + \beta \frac{\partial J}{\partial \hat{\lambda}(r;\theta)^{(\nu)}} \tag{4.55}$$

where ν is the iteration number.

Equations (4.54) and (4.55) are the basis of the reconstruction algorithm illustrated in this section. For a steepest ascent algorithm, the step length β is usually obtained by a one-dimensional search. For example, Wernecke and D'Addario (1977) use the following approach: take a step of predetermined length in the search direction, evaluate the

function and gradient there, and estimate the location of the one-dimensional search maximum by quadratic or cubic interpolation. The approach followed here, however, is to use a fixed step length β.

Care is needed over the partial derivative on the right side of (4.54) when the form of $\Phi\{\cdot\}$ is such (as it is for two of the methods examined in this section) that there is a hiatus where $\hat{\lambda}(r;\theta)$ passes through zero. Following Wernecke and D'Addario (1977), the approach adopted here is to divert the search direction near zero so that the estimated density cannot fall below some prescribed real and non-negative constant which is denoted by α. Thus

$$\hat{\lambda}(r;\theta) \geq \alpha \geq 0 \qquad \text{for all } (r;\theta) \in \Omega \tag{4.56}$$

is called the *non-linear step* in the iterative image reconstruction algorithm.

An initial guess for $\hat{\lambda}(r;\theta)$ is needed to start the iterations. Experience has shown that final convergence is not critically dependent upon the "goodness" of this guess, although the rate of convergence increases the closer the initial $\hat{\lambda}(r;\theta)$ is to the actual $\lambda(r;\theta)$. Here the initial guess is taken to be the result of modifying and back-projecting the given $\{f,\phi_n,N\}$. The form of the right side of (4.54) is such that the maximisation of J can be reduced to a straightforward iterative scheme, which it should be noted is similar to the SIRT algorithm (cf. Gilbert, 1972):

(i) Compute $\partial\Phi\{\cdot\} / \partial\hat{\lambda}(r;\theta)$ at all points in the x,y-grid (as defined in §4.2.1).

(ii) Compute $\hat{f}(\xi;\phi)$, at each data point (as defined in §4.2.1).

(iii) Compute $f_d(\xi;\phi)$, as defined by (4.45), at each data point.

(iv) Compute $\lambda_d(r;\theta)$, as defined by (4.53), at each point on the x,y-grid.

(v) Cease iterations when, at all points on the x,y-grid, $\lambda_d(r;\theta)$ is below whatever threshold is deemed appropriate for the envisaged application.

(vi) Compute $\partial J / \partial\hat{\lambda}(r;\theta)$, as defined by (4.54).

(vii) Compute the next estimate for $\hat{\lambda}(r;\theta)$ from (4.55).

(viii) Continue until the iterations cease at step (v).

§§4.5.1 through 4.5.3 illustrate the use of this scheme with four reconstruction methods (V - VIII). It is worth emphasising that these

schemes have been developed for the problem of reconstructing from few projections. They can be considered to interpolate between the given projections *implicitly*.

4.5.1 Methods V and VI: Maximum entropy

Two forms of "entropy" are employed with roughly equal frequency in the literature. They are discussed separately below.

Method V: The functional $\Phi\{\cdot\}$ is the entropy 1 expression (from Table 4.5). Thus

$$\frac{\partial\Phi\{\hat{\lambda}(r;\theta)\}}{\partial\hat{\lambda}(r;\theta)} = \frac{1}{\hat{\lambda}(r;\theta)} \qquad (4.57)$$

which is required in step (vi) of the iterative scheme. This derivative is undefined for $\hat{\lambda}(r;\theta) = 0$ so the non-linear step (4.56) is added to the iterative scheme after step (vii). Figure 4.16 shows cuts through the reconstruction of the test phantom using method V. The figure caption gives the parameters used to obtain these results.

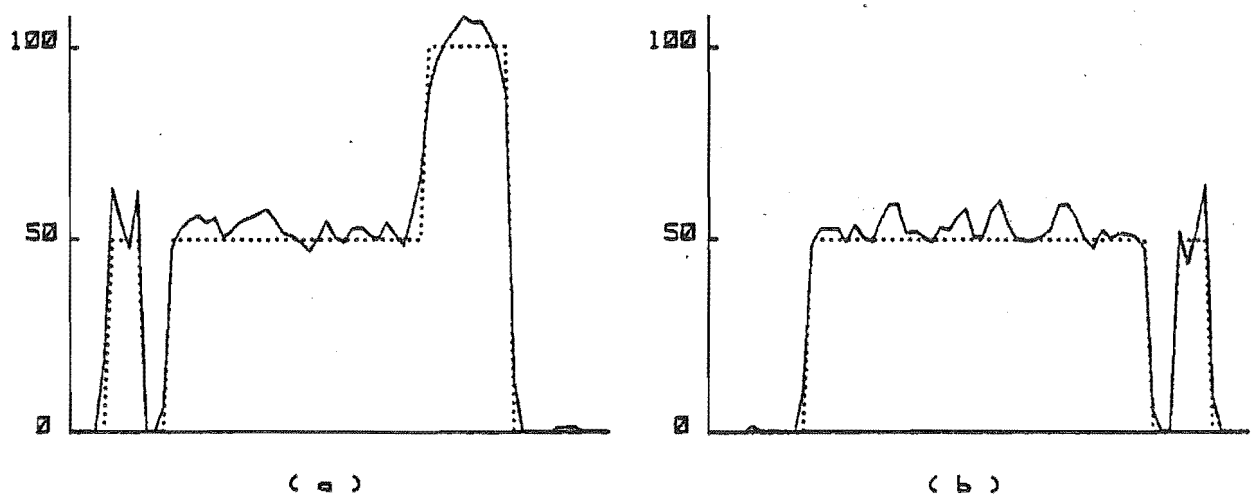


Figure 4.16 Density cuts through the test phantom (•••) and its reconstruction by method V (——): (a) cut X - X, (b) cut Y - Y. The following parameters (defined in the text) apply: $N = 8$, $\alpha = 0.1$, $\beta = 0.05$, $\gamma = 0.25$.

Method VI: The second entropy expression from Table 4.5 is used for the functional $\Phi\{\cdot\}$. For this method

$$\frac{\partial\Phi\{\hat{\lambda}(r;\theta)\}}{\partial\hat{\lambda}(r;\theta)} = -1 - \ln\left(\hat{\lambda}(r;\theta)\right). \qquad (4.58)$$

The non-linear step (4.56) is again required to avoid the reconstructed density going to zero. Figure 4.17 shows cuts through the reconstructed

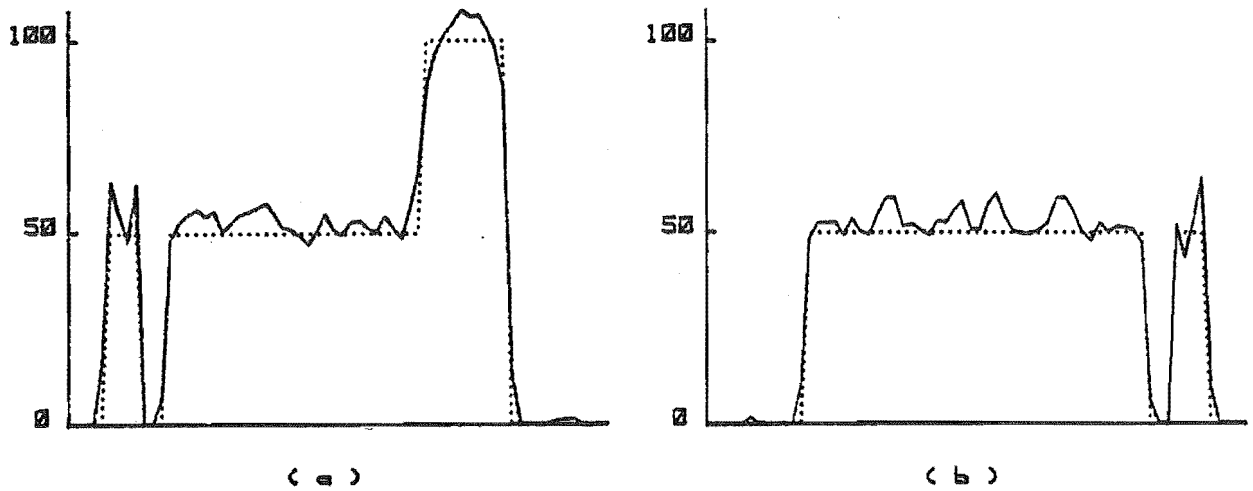


Figure 4.17 Density cuts through the test phantom (•••) and its reconstruction by method VI (——): (a) cut X - X, (b) cut Y - Y. The following parameters apply: N = 8, α = 0.0001, β = 0.05, γ = 0.25.

density. The results obtained from methods V and VI are very similar. The two entropy expressions are based on different models of the image statistics, yet they behave similarly when they are included in the above iterative scheme. This has suggested that it might be worthwhile investigating other methods based on the above iteration scheme, but which employ functionals of simpler form. Two such methods are discussed in §§4.5.2 and 4.5.3.

4.5.2 Method VII: Positivity constraint

For method VII the functional $\Phi\{\cdot\}$ is chosen to be a constant. Thus

$$\frac{\partial \Phi\{\hat{\lambda}(r;\theta)\}}{\partial \hat{\lambda}(r;\theta)} = 0 \qquad (4.59)$$

and the iteration scheme (4.54 and 4.55) becomes

$$\hat{\lambda}(r;\theta)^{(\nu+1)} = \hat{\lambda}(r;\theta)^{(\nu)} - \beta\,\gamma\,\lambda_d(r;\theta) \qquad (4.60)$$

where the difference projections used to calculate $\lambda_d(r;\theta)$ are obtained from the current estimate of the image $\hat{\lambda}(r;\theta)^{(\nu)}$. The step length β and the multiplier γ have become one parameter. For comparison with methods V and VI the non-linear step is included in the iterative scheme. When $\alpha = 0$ this method is the same as the SIRT algorithm (Gilbert, 1972). Figure 4.18 shows cuts through the image of the test phantom reconstructed by this method.

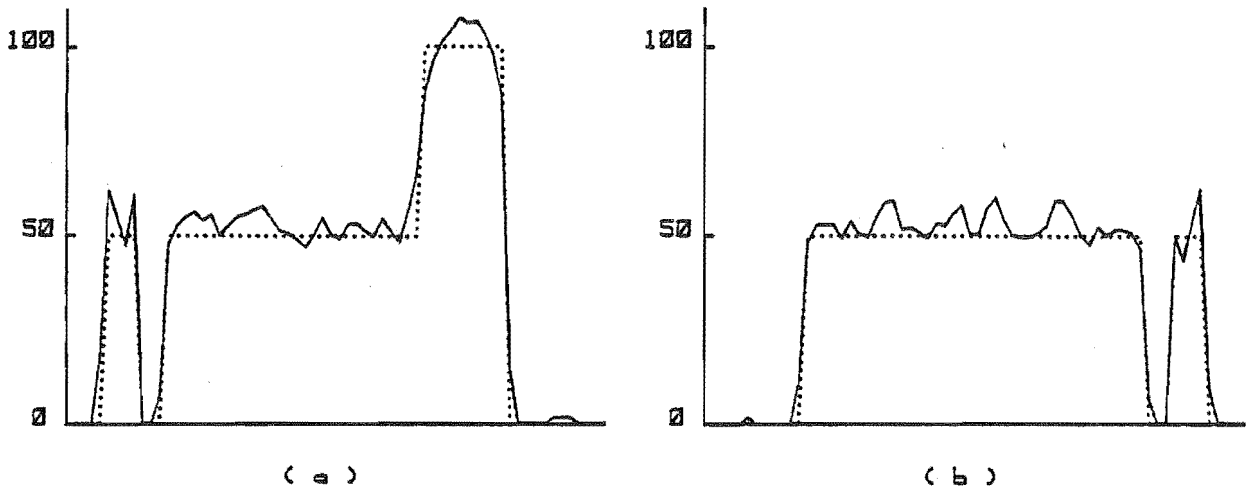


Figure 4.18 Density cuts through the test phantom (•••) and its reconstruction by method VII (——): (a) cut X - X, (b) cut Y - Y. The following parameters apply: $N = 8$, $\alpha = 0$, $\beta\gamma = 0.025$.

4.5.3 Method VIII: Size constraint

Method VIII is similar to method VII, except that at the non-linear step the image is set to zero outside its known extent. For the example presented here, the extent of the image is specified by the mask defined in §4.2.8. Thus methods VII and VIII are analogous to the two Gerchberg inspired methods (III and IV) in that one of each (III and VII) uses only *a priori* knowledge of image positivity, and the second (IV and VIII) uses image positivity and extent. The difference between the methods is in the way in which the schemes operate. Figure 4.19 shows the reconstruction of the test phantom obtained by method VIII.

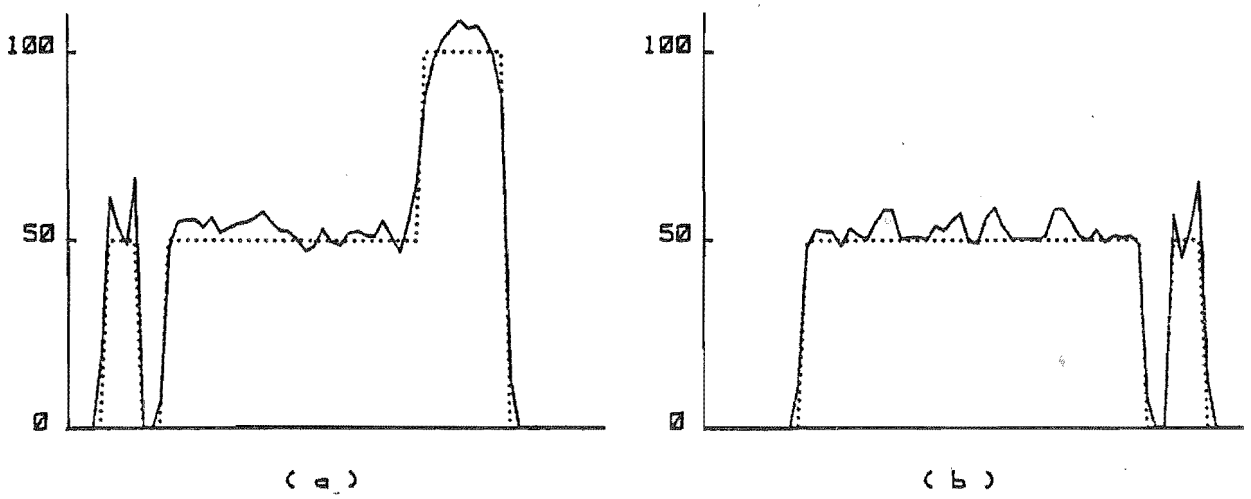


Figure 4.19 Density cuts through the test phantom (•••) and its reconstruction by method VIII (——): (a) cut X - X, (b) cut Y - Y. The following parameters apply: $N = 8$, $\alpha = 0$, $\beta\gamma = 0.025$.

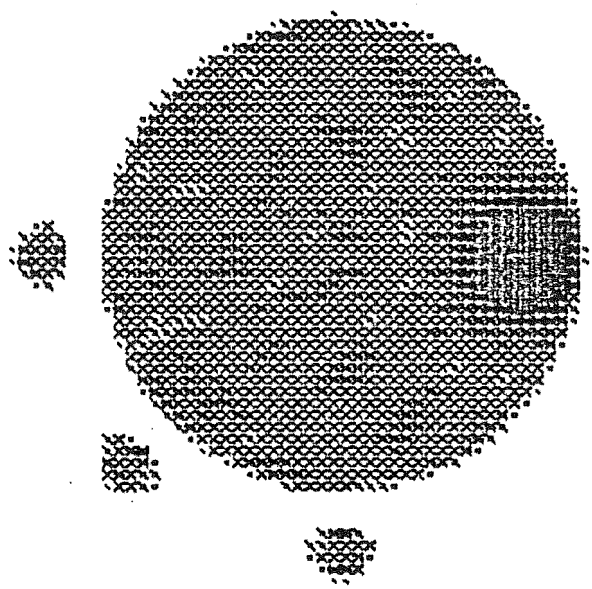


Figure 4.20 Density plot of the reconstruction obtained by method VIII. The parameters listed in the caption for Figure 4.19 apply. The grey-scale covers the range of 0 (white) to 100 (black).

Figure 4.20 is a density plot of this image which should be compared with Figures 4.1, 4.2, 4.4 and 4.14. Table 4.6 summarises the results obtained by methods V through VIII.

Table 4.6 Comparison criteria for methods V-VIII (the symbol T stands for the Roman numeral identifying any particular method) applied to the test phantom. There are eight given projections.

T	α	β ; γ	Number of iterations	$\varepsilon(T,\Xi)$	$\varepsilon(T,\Lambda)$
V	0.1	$\beta = 0.05$; $\gamma = 0.25$	233	.35	.61
VI	0.0001	$\beta = 0.05$; $\gamma = 0.25$	249	.34	.61
VII	0	$\beta\gamma = 0.025$	122	.35	.61
VIII	0	$\beta\gamma = 0.025$	89	.34	.60

The number of iterations performed in each of the examples is included in Table 4.6. The convergence rate for method VIII is illustrated in Figure 4.21. This example was continued until there was no further image improvement. However, this does not mean that the difference density $\lambda_d(r;\theta)$ has gone to zero (cf. (4.60)).

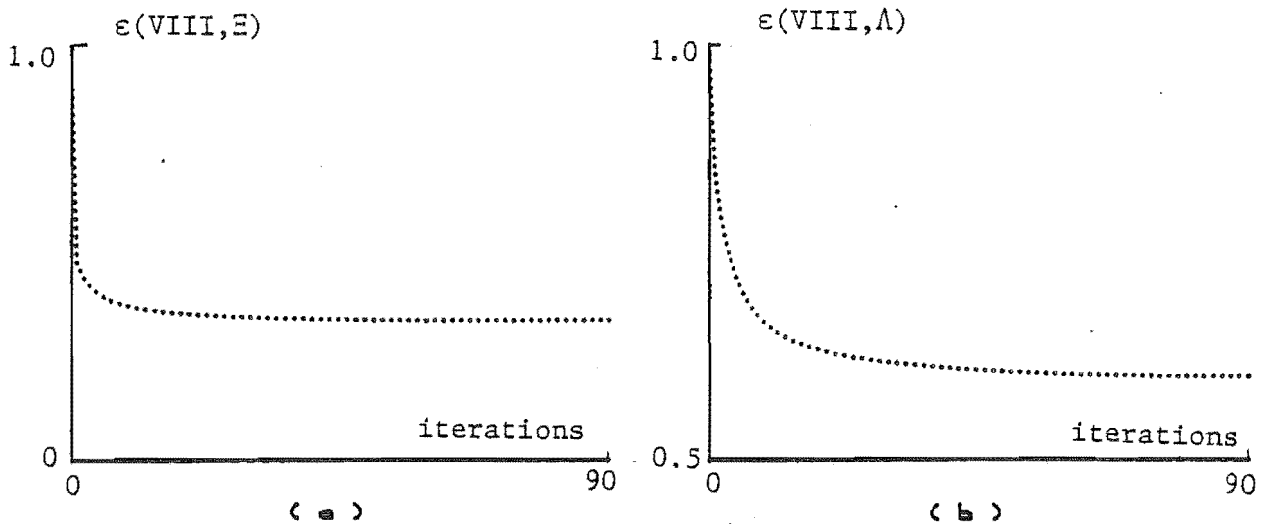


Figure 4.21 Convergence rates for method VIII when applied to the test phantom: (a) the error criterion for the whole of image space, (b) the error criterion for the region actually occupied by the cross-section of the object. The parameters listed in the caption for Figure 4.19 apply.

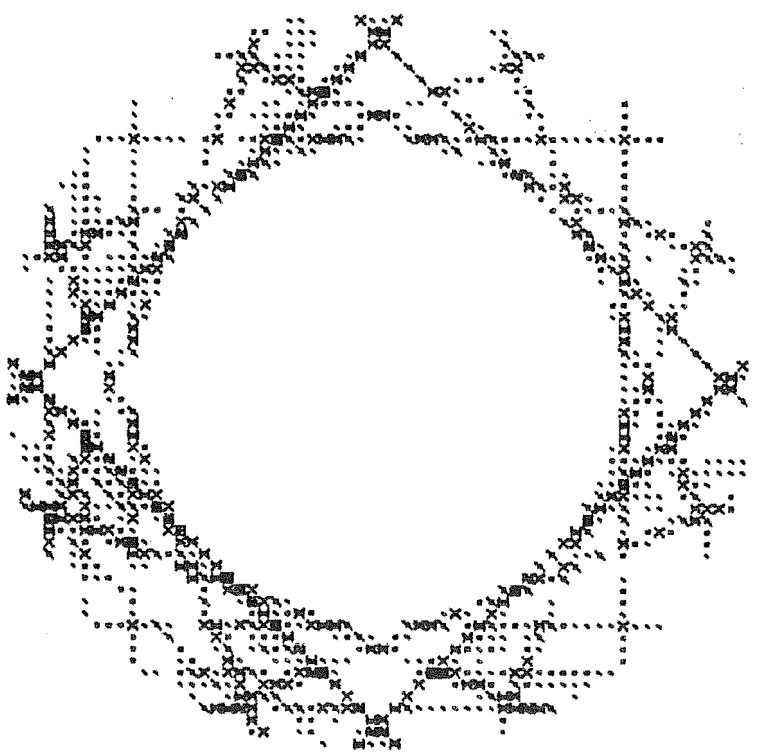


Figure 4.22 The difference density remaining after 89 iterations when method VIII ceased to converge towards the test phantom. The parameters listed in the caption for Figure 4.19 apply. The grey-scale covers the range of 0 (white) to 10 (black).

The difference density (4.53) depends on $\hat{f}(\xi;\phi)$ - the projections calculated from the current estimate of the image. In §4.2 it is shown how interpolation errors in the re-projection scheme cause image degradation. The re-projection method used here employs Peters' (1981) correction scheme, but Table 4.2 shows how even this scheme causes about 0.3% error for one iteration when applied to a gaussian density distribution. Thus the iteration schemes described above cease to improve $\hat{\lambda}(r;\theta)$ once the errors introduced at that iteration equal the improvement which could be given by an error-free scheme. Figure 4.22 shows the difference density after 89 iterations, which was when the estimated image provided by method VIII ceased converging on the true image. As would be expected, the difference density is lowest in the central region where the image is relatively constant. The largest errors are streaks, aligned at the projection angles, and emanating from the positions of the edges of the discs in the object. This is because the interpolation errors introduced by the re-projection scheme are largest at sharp transitions in density.

4.6 DISCUSSION

This chapter has examined eight methods for reconstructing images from few projections. Methods which use *a priori* image information give better images than those obtained from straightforward reconstruction from the measured projections or when intermediate projections are first interpolated by a scheme such as FSI (method II). Tables 4.3, 4.4 and 4.6 summarise the results obtained by the methods developed in §§4.3 and 4.5. Comparison of Table 4.1 with Tables 4.3, 4.4 and 4.6 shows the kind of improvement that can be afforded by each of the methods. Here, the error criterion $\varepsilon(T,\Lambda)$ is the important way of comparing the methods. Thus when there are eight given projections, method III gives a best image which has an error equivalent to reconstruction from 14 given projections by method I. By the same argument method IV gives an improvement equivalent to 14 given projections, and method VIII gives an improvement equivalent to 16 given projections.

Methods V-VIII are seen to give better results than the methods based on Gerchberg's algorithm (III and IV). One particular advantage of the iteration scheme developed in §4.5 is that each estimate of the image is updated by the difference projections, which are the difference

between the measured projections and those calculated from the current estimate of the image. Thus any errors in $\hat{f}(\xi;\phi)$ tend to be corrected at each iteration. However, in Gerchberg's algorithm (§4.3) each estimate of the image is formed from the measured projections plus those estimated at intermediate angles from the current estimate of the image. Errors in those projections computed at the intermediate angles cause errors in the next estimate of the image. Projections are subsequently recalculated from this new estimate. These are degraded by errors in the image plus any errors introduced by the re-projection scheme. There is no correction to these projections at the intermediate angles and so errors build up in the projections at intermediate angles with successive iterations (cf. §4.2.6). This explains why Gerchberg's algorithm gives worse results than the schemes introduced in §4.5. Since errors do not propagate in those schemes, the accuracy of their final images is not degraded by the large number of iterations.

The results presented in §4.5 show that methods V-VIII are useful reconstruction schemes, but too slow to be of routine practical use. The optimisation scheme used there is one which employs a fixed step length but which could easily be accelerated by one of the algorithms discussed in §4.4. Considerable improvement might be possible. For instance, Trussell and Hunt (1979) report that using a modified Picard iteration scheme rather than steepest ascent results in a factor of ten less iterations for convergence, and Jain and Ranganath (1981) report a similar improvement with a conjugate gradient algorithm.

Comparison of methods V through VIII shows that it is the general approach that is important, rather than the detailed form of the functional $\Phi\{\cdot\}$. It does not seem to matter what kind of entropy one adopts, or whether one employs a different functional. In fact, the results obtained by methods VII and VIII, which replace the functional with a constant, are (perhaps slightly) superior. This is an interesting conclusion in view of the arguments postulated for statistically-based functionals (cf. Frieden, 1980). Several authors (cf. Högbom, 1979; Bracewell, 1980) have recently questioned the use of entropy-based functionals for image restoration.

In conclusion, the choice of a reconstruction method for a particular application is a trade-off between accuracy and speed. Thus method I would be used when the accuracy is sufficient for the particular

application. Schemes such as method VIII would be used when there is a small number of projections and the added accuracy is important. Before such a scheme could be used routinely it would have to be modified, as suggested above, to accelerate its convergence. The practical significance of this is discussed in §5.3.

5. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

5.1 COMPUTER AIDED LEARNING

The first application of modelling studied here is one in which the model forms the basis of a teaching program. The model takes a simple view of halothane anaesthesia, yet allows quite realistic simulations for teaching purposes. The program has been evaluated in a pilot study. The evaluation, which is presented in §2.6, is based on a questionnaire which was given to a group of students before, and again after, a session using the program. The results show that the program significantly aids learning. There were insufficient students to test the program in comparison to other ways of teaching the same material (e.g. a lecture). This is something that could be tried if greater numbers of students became available.

The program falls into the third category of CAL programs discussed in §1.4 - it embodies a dynamic simulation, yet is not intended to teach the skills required to administer halothane. The aim has been to reinforce the principles of anaesthetic uptake and distribution. The development of this program suffered from the difficulties associated both with the design and implementation of an interactive program and with the formulation of a model.

The first difficulty, and one of the reasons for the slow acceptance of CAL, is the effort involved in producing a program. Development of software is increasingly being simplified by the use of high-level programming languages. There are aids such as author languages for the development of multi-choice and branched-tree structure type programs. These sorts of problems can be expressed in a relatively standard format. There is a need for an author language which is more suited to the development of the third type of CAL program. Such simulation-based programs are more difficult to constrain to a standard format.

The second difficulty associated with developing a CAL simulation program is in defining the model. There is no general method for this. Each application has its own particular problems. Here, the model was constructed from the data which are given in the literature. Gaps and

inconsistencies in this data were resolved with the aid of the clinical experience of members of the staff of the Department of Anaesthetics, Christchurch Clinical School. The development of this model has isolated areas where current knowledge is weak. This has meant that there are several limitations of the model. For example, it does not include the effects of CO_2 or the effects of other drugs. Also the effects of halothane on heart rate, ventilation and stroke volume are modelled somewhat empirically. There has been little choice in this regard however, since the available data are so sparse. Thus there is a need for more research into halothane anaesthesia. If there were more data available then a more clinically realistic simulation could be designed. Despite the advantages of a program of the third class, there is also a need for a program which is realistic enough to allow skill-learning.

Improvements to the student-computer interface are among the major advances that can be expected in the near future. Simulation programs especially will benefit from the availability of better terminals, particularly those capable of colour graphics. For many institutions the prices of such terminals are currently prohibitive, but there is no doubt that the prices will soon become competitive. Other hardware developments such as the increasing availability of small, micro-processor based computers are also having an impact on CAL. However the speed and storage requirements of sophisticated simulation programs make them difficult to implement on small computers. Advances such as free-format dialogue and speech recognition are also beginning to be used in CAL. One of the disadvantages of most CAL programs currently used is that they require the student to use a keyboard. Thus the capability of speech communication with the computer will greatly simplify the interface between student and computer and open up new areas for CAL.

5.2 PATIENT MOVEMENT IN CT

Chapter 3 illustrates another application of modelling. The aim there is to characterise the blurring due to patient movement in computed tomography. The approach taken is to consider an object which consists of one point which is allowed to move in different ways while the projections are taken. Several different types of movement of the object point are investigated. Each type of movement is a model for motion that

could occur in practice (providing it is in the plane of interest). When motion has occurred, the reconstructed image is a blurred representation of the object point. The radial dependence of this blurring is quantified in Chapter 3. Of course, in practice, the form of any motion that may have occurred is rarely known. Thus in Chapter 3, statistical averages are taken over classes of movement so that quite general conclusions are obtained. It is shown that the resolution limit set by motion which occurs for the whole of the time of the scan, is generally about twice the effective radius of the motion. When motion occurs for only part of the time of the scan, the blurring can sometimes be negligible.

These results characterise motion in the plane. Motion perpendicular to the plane of interest has not been considered. However, computational experience shows that the artefacts due to this "z-axis" motion are less severe than those studied here (cf. Alfidi *et al.*, 1976). The reasons for this need more study.

The approach taken here has been only to characterise the blurring due to motion. Because the motion is generally unknown it is usually not possible to correct for it. However, a human observer can usually recognise an image even though it has streaks superimposed on it. The reason for this is that one has significant *a priori* knowledge of the type of structures that are in an image. For example, in an image of the heart one does not expect to see any straight lines. Thus experience and knowledge of anatomy often helps an observer to "see through" or ignore artefacts in the image. This suggests that artefacts may be able to be reduced if *a priori* knowledge is built into the reconstruction algorithm. Thus the incorporation of some sort of a model of the anatomy may be a way of overcoming the artefacts due to patient motion.

5.3 IMAGE RECONSTRUCTION FROM FEW PROJECTIONS

Several methods for reconstructing images from few projections are investigated in Chapter 4. It is found that methods which employ models of the image give reconstructions which are superior to those obtained by straightforward procedures such as modified back-projection. The models are based on the *a priori* knowledge which is available. For example it is usually known that the image is real and non-negative. An estimate of the extent of the image may also be known.

The two methods investigated in §4.3 are based on Gerchberg's (1974) algorithm. The numerical properties of such algorithms are studied by Papoulis (1975) who examines the errors introduced by truncation of the iterations, noise, and computer roundoff. Stark *et al.* (1981 b,c) have recently generalised the underlying theory of such algorithms (cf. Youla, 1978) in order to establish the conditions in which a unique reconstruction is possible. Gerchberg's algorithm has been used extensively in CT when there is a missing sector of projection angles (cf. Sato *et al.*, 1981; Tam and Perez-Mendez, 1981). However, in Chapter 4 it is shown that this scheme is sensitive to numerical errors when it is used to reconstruct images from few projections and this limits the number of iterations that can be done before the image starts to diverge. If Gerchberg's algorithm is to produce higher quality images than those obtained in §4.3, numerical routines of greater accuracy than those developed in §4.2 are needed.

The four reconstruction methods investigated in §4.5 are not as sensitive to numerical errors as the methods based on Gerchberg's algorithm. The first two of these are based on maximising the entropy of the image. However it was found that the methods worked better when the entropy maximisation was replaced with a simple positivity constraint. This accords with the conclusions of Tam and Perez-Mendez (1981) who report that several images which have different r.m.s. error, all have the same entropy (that of the original object). Thus entropy is an insensitive measure of image quality.

In Chapter 4 it is assumed that there is sufficient ξ-sampling - but only small numbers of projections. Shim and Cho (1981) consider reconstruction when there are few given projections and when there are only a small number of samples in each projection. They present a singular value decomposition (SVD) pseudoinversion method and report that it gives superior reconstructions to those given by modified back-projection. Thus it would be useful to compare the SVD pseudoinversion method with the methods presented in §4.5 in the case when there are few given projections and also when the sampling of each projection is coarse.

Unfortunately, although the methods developed in §4.5 give the most accurate reconstructions, they are the slowest of the methods investigated. However the convergence rate can easily be improved by one of the schemes reviewed in §4.4. This is an important avenue for future research.

A particular case of interest in which there is only a small number of measured projections occurs when imaging the beating heart. The technique of McKinnon and Bates (1981) produces small numbers of projections for each phase of the heart's motion using a conventional CT scanner. When used in its fastest mode, the DSR (Robb *et al.*, 1979) also produces only small sets of projections for each phase of the heart cycle. In such cases it is desirable to make the best use of the small set of available projections. The schemes investigated in §4.5 show that it is possible to obtain reconstructions which are superior to those obtained by straightforward modified back-projection. However in applications where there are many images to be reconstructed (as with the DSR) there is currently no alternative to using modified back-projection since it is so much faster than iterative reconstruction methods. Despite the speed requirements of producing serial images of the beating heart, this application provides the opportunity for the extensive use of *a priori* information. Consider that there exists an image of the heart at one particular instant. It is required to estimate a new image of the same cross-section a short time later. There may be only a small set of projections available at this new time. These projections can be used to form a new image independent of all other images. However the attractive feature of this situation is that a great deal is already known about the cross-section from the previous images, and it is known that there are only small changes from one image to the next, i.e. there is considerable correlation between one image and the one at the next instant. What is needed is an adaptive image reconstruction method. This is an interesting area for future research.

APPENDIX 1

CAL EVALUATION QUESTIONNAIRE

(The correct answers to the first four questions are marked with an asterisk.)

1. At the end of 20 minutes of halothane anaesthesia (>1% inspired concentration), the concentration of halothane (expressed in volumes per cent) is highest in:
   1. Arterial blood*
   2. Venous blood
   3. Fat
   4. Brain
   5. Muscle

2. After 20 minutes of halothane anaesthesia, the vaporiser is turned off. Which one of the following is true:
   1. The concentration of halothane in arterial blood will not fall significantly for some minutes.
   2. The concentration of halothane in muscle does not fall for some minutes.*
   3. Elimination of halothane from the body occurs at a constant rate.
   4. A fall in cardiac output will hasten the rate of elimination of halothane.
   5. None of the above.

3. After two hours of halothane anaesthesia, the concentration of halothane in the brain depends mainly on:
   1. Cardiac output
   2. Alveolar ventilation
   3. Brain mass
   4. Cerebral venous halothane concentration
   5. Alveolar halothane concentration*

4. After 20 minutes of halothane anaesthesia, which one of the following is correct:
   1. There has been an increase in heart rate
   2. Blood pressure has not changed
   3. The pulse pressure has decreased*
   4. Cardiac output has tended to increase
   5. None of the above

Questions 5-8 (identical to questions 1-4) are not reproduced here.

9. The program displays information in a graphical form on the terminal. The horizontal axis of the graph represents:
   1. Inspired halothane fraction
   2. Time
   3. Brain halothane concentration
   4. Inspired gas flow rate

10. The program presents tissue concentrations of halothane as:
   1. moles per litre
   2. gm per 100 ml
   3. litres (STPD) per 100 ml
   4. volumes per cent.

11. This computer program is not capable of accurately predicting what you would observe during practical anaesthesia. Indicate which of the following are the <u>two</u> most important reasons for this:
   1. No account is taken of the biological variation which occurs in the real world.
   2. The graphs on the terminal are too coarse to be read with sufficient accuracy.
   3. The settings on the knobs controlling halothane fraction and fresh gas flow are not accurate.
   4. Some of the physiological interactions which affect halothane uptake and distribution are not well understood and cannot therefore be represented by a computer program.
   5. The method used by the computer program to solve the equations which describe the uptake and distribution of halothane is only approximate.

12. Have you used this program before?
    1. Yes
    2. No

13. Do you think this tutorial was:
    1. Helpful and relevant to your course
    2. Not relevant but interesting
    3. Too difficult to understand
    4. A waste of time.

14. This program is available on the CAL terminal whenever the library is open. Would you like to:
    1. Return and use the program by yourself
    2. Attend more tutorial sessions based on this program
    3. Have nothing more to do with this program.

REFERENCES

Abramowitz M. and Stegun I.A. 1965. "Handbook of Mathematical Functions", Dover Publications, New York.

Adby P.R. and Dempster M.A.H. 1974. "Introduction to Optimization Methods", Chapman and Hall, London.

Adkins G. and Pooch U.W. 1977. "Computer Simulation: A Tutorial", Computer 10, April 12-17.

Alfidi R.J., MacIntyre W.J., and Haaga J.R. 1976. "The Effects of Biological Motion on CT Resolution", Am J Roentgenol 127, 11-15.

Alvarez R.E. and Macovski A. 1976. "Energy-selective Reconstructions in X-ray Computerized Tomography", Phys Med Biol 21, 733-744.

Andrews H.C. and Hunt B.R. 1977. "Digital Image Restoration", Prentice-Hall, New Jersey.

Baba N., Murata K., Okada K., and Fujimoto Y. 1981. "Image reconstruction from electron micrographs by maximum entropy method", Optik 58, 233-239.

Bagshaw R.J. 1980. "Systems Theory and the Anaesthetist", Acta Anaesth Scand 24, 379-392.

Bates J.H.T. 1981. "Applications of Modelling and Image Processing in Medicine", Ph.D. thesis, University of Otago, Dunedin, New Zealand.

Bates R.H.T. and Peters T.M. 1971. "Towards Improvements in Tomography", NZ J Sci 14, 883-896.

Bergland G.D. 1969. "A guided tour of the fast Fourier transform", IEEE Spectrum 6, 41-52.

Bracewell R.N. 1977. "Correction for Collimator Width (Restoration) in Reconstructive X-Ray Tomography", J Comput Assist Tomogr 1, 6-15.

Bracewell R.N. 1980. Review of "Image Formation from Coherence Functions in Astronomy (1979) ed. by C. van Schooneveld and publ. by the D. Reidel Publishing Co., Dordecht, Holland" in J Opt Soc Am 70, 570.

Braun L., Finkelstein S.M., and Blesser W.B. 1977. "Pulmonary Diffusion - A Model Study", Computer 10, 28-36.

Brooks R.A. and Di Chiro G. 1976. "Principles of Computer Assisted Tomography (CAT) in Radiographic and Radioisotopic Imaging", Phys Med Biol 21, 689-732.

Brooks R.A., Weiss G.H., and Talbert A.J. 1978. "A New Approach to Interpolation in Computed Tomography", J Comput Assist Tomogr 2, 577-585.

Chambers J.A. and Sprecher J.W. 1980. "Computer Assisted Instruction: Current Critical Issues", Commun ACM 23, 332-342.

Chambers W.S. 1977. "AWAVS: An Engineering Simulator for Design of Visual Flight Training Simulators", J Aircr 14, 1060-1063.

Chattergy R. and Pooch U.W. 1977. "Integrated Design and Verification of Simulation Programs", Computer 10, 40-45.

Cormack A.M. 1978. "Sampling the Radon Transform with Beams of Finite Width", Phys Med Biol 23, 1141-1148.

Cowles A.L., Borgstedt H.H., and Gillies A.J. 1971. "Tissue Weights and Rates of Blood Flow in Man for the Prediction of Anesthetic Uptake and Distribution", Anesthesiology 35, 523-526.

Deutsch S., Linde H.W., Dripps R.D., and Price H.L. 1962. "Circulatory and Respiratory Actions of Halothane in Normal Man", Anesthesiology 23, 631-638.

Dickinson C.J. 1977. "A computer model of human respiration: Ventilation, blood gas transport and exchange, hydrogen ion regulation ... 'MacPuf'", MTP Press, Lancaster.

Duerinckx A.J. and Macovski A. 1978. "Polychromatic Streak Artifacts in Computed Tomography Images", J Comput Assist Tomogr 2, 481-487.

Duerinckx A.J. and Macovski A. 1979. "Classification of Artifacts in X-ray CT Images Due to Non-linear Shadows", IEEE Trans NS-26, 2848-2852.

Durrani T.S. and Goutis C.E. 1980. "Optimisation techniques for digital image reconstruction from their projections", Proc IEE 127(E), 161-169.

Eger E.I., Smith N.T., Cullen D.J., Cullen B.F., and Gregory G.A. 1971. "A Comparison of the Cardiovascular Effects of Halothane, Fluroxene, Ether and Cyclopropane in Man: A resumé", Anesthesiology 34, 25-41.

Eger E.I. 1974. "Anesthetic Uptake and Action", Williams and Wilkins, Maryland.

Elgerd O.I. 1967. "Control Systems Theory, McGraw-Hill, Tokyo.

Frieden B.R. 1972. "Restoring with Maximum Likelihood and Maximum Entropy", J Opt Soc Am 62, 511-518.

Frieden B.R. 1980. "Statistical Models for the Image Restoration Problem", Computer Graphics and Image Processing 12, 40-59.

Gerchberg R.W. 1974. "Super-resolution through error energy reduction", Optica Acta 21, 709-720.

Gilbert P. 1972. "Iterative Methods for the Three-dimensional Reconstruction of an Object from Projections", J theor Biol 36, 105-117.

Glover G.H. and Eisner R.L. 1979. "Theoretical Resolution of Computed Tomography Systems", J Comput Assist Tomogr 3, 85-91.

Glover G.H. and Eisner R.L. 1980. "Correction to Theoretical Resolution of Computed Tomography Systems", J Comput Assist Tomogr 4, 108.

Gore J.C. and Leeman S. 1980. "The reconstruction of objects from incomplete projections", Phys Med Biol 25, 129-136.

Gull S.F. and Daniell G.J. 1978. "Image reconstruction from incomplete and noisy data", Nature 272, 686-690.

Guyton A.C. 1971. "Textbook of Medical Physiology", 4th edn, W.B. Saunders Co., Philadelphia.

Hansen E.W. 1981a. "Theory of circular harmonic image reconstruction", J Opt Soc Am 71, 304-308.

Hansen E.W. 1981b. "Circular harmonic image reconstruction: experiments", Appl Optics 20, 2266-2274.

Harding G. 1981. "X-ray Imaging with Scattered Radiation", submitted to J Sci Instr (J Phys E).

Herman G.T., Lent A., and Rowland S.W. 1973. "ART: Mathematics and Applications - A report on the Mathematical Foundations and on the Applicability to Real Data of the Algebraic Reconstruction Techniques", J theor Biol 42, 1-32.

Herman G.T. 1979. "Correction for Beam Hardening in Computed Tomography", Phys Med Biol 24, 81-106.

Heuscher D.J. and Carlson R.W. 1979. "Partial Area Reconstruction of the Spine Using a Projection Completion Method", Presented at the 65th Scientific Assembly, Georgia World Congress Center, Atlanta, Georgia, Nov 25-30, 1979.

Hinderling Th., Rüegsegger P., Anliker M., and Dietschi C. 1979. "Computed Tomography Reconstruction from Hollow Projections: An Application to *In Vivo* Evaluation of Artificial Hip Joints", J Comput Assist Tomogr 3, 52-57.

Hoffer E.P., Barnett G.O., Farquhar B.B., and Prather P.A. 1975. "Computer-aided Instruction in Medicine", Annual Review of Biophysics and Bioengineering 4, 103-118.

Högbom J.A. 1979. "The Introduction of *a priori* Knowledge in Certain Processing Algorithms", in "Image Formation from Coherence Functions in Astronomy", edited by C. van Schooneveld, publ. by the D. Reidel Publishing Co., Dordecht, Holland, pp. 237-239.

Huesman R.H. 1977. "The Effects of a Finite Number of Projection Angles and Finite Lateral Sampling of Projections on the Propagation of Statistical Errors in Transverse Section Reconstruction", Phys Med Biol 22, 511-521.

Inouye T. 1979. "Image Reconstruction with Limited Angle Projection Data", IEEE Trans NS-26, 2666-2669.

Isaacson E. and Keller H.B. 1966. "Analysis of Numerical Methods", John Wiley and Sons, New York.

Jain A.K. 1981. "Advances in Mathematical Models for Image Processing", Proc IEEE 69, 502-528.

Jain A.K. and Ranganath S. 1981. "Extrapolation Algorithms for Discrete Signals with Application in Spectral Estimation", IEEE Trans ASSP-29, 830-845.

Joseph P.M. and Spital R.D. 1978. "A method for Correcting Bone Induced Artifacts in Computed Tomography Scanners", J Comput Assist Tomogr 2, 100-108.

Kenny G.N.C. and Schmulian C. 1979. "Computer-assisted learning in the teaching of anaesthesia", Anaesthesia 34, 159-162.

Kowalski G. 1977a. "Influence of Fixed Errors of a Detector Array on the Reconstruction of Objects from their Projections", IEEE Trans NS-24, 2006-2016.

Kowalski G. 1977b. "Reduction of Circular Artifacts in Fan-Beam Scanners", IEEE Trans NS-24, 2555-2560.

Kowalski G. 1977c. "Reconstruction of Objects from their Projections. The Influence of Measurement Errors on the Reconstruction", IEEE Trans NS-24, 850-864.

Kowalski G. and Wagner W. 1977. "Generation of pictures by X-ray scanners", Optica Acta 24, 327-348.

Lamb M.R. and Bates R.H.T. 1978. "Computerized Aural Training: An Interactive System Designed to Help Both Teachers and Students", J Comput Based Instr 5, 30-37.

Lewitt R.M. and Bates R.H.T. 1978a. "Image reconstruction from projections. I: General theoretical considerations", Optik 50, 19-33.

Lewitt R.M. and Bates R.H.T. 1978b. "Image reconstruction from projections. III: Projection completion methods (theory)", Optik 50, 189-204.

Lewitt R.M. and Bates R.H.T. 1978c. "Image reconstruction from projections. IV: Projection completion methods (computational examples)", Optik 50, 269-278.

Lewitt R.M., Bates R.H.T., and Peters T.M. 1978. "Image reconstruction from projections. II: Modified back-projection methods", Optik 50, 85-109.

Lowe H.J. 1972. "Dose-Regulated Penthrane Anesthesia", Abbott Laboratories, Chicago.

McIntyre J.W.R. 1980. "Computer-Aided Instruction as Part of an Undergraduate Programme in Anaesthesia", Can Anaesth Soc J 27, 68-73.

McKinnon A.E. 1973. "Computing Calcium Dynamics in Man", Ph.D. thesis, University of Canterbury, Christchurch, New Zealand.

McKinnon G.C. and Bates R.H.T. 1981. "Towards Imaging the Beating Heart Usefully with a Conventional CT Scanner", IEEE Trans BME-28, 123-127.

Mapleson W.W. 1973. "Circulation-time Models of the Uptake of Inhaled Anaesthetics and Data for Quantifying them", Br J Anaesth 45, 319-334.

Mapleson W.W. 1979. "From Glover to Computer. Towards programmed anaesthesia?", Anaesthesia 34, 163-172.

Miller L.A. and Thomas J.C. 1977. "Behavioral issues in the use of interactive systems", Int J Man-Machine Studies 9, 509-536.

Minerbo G. 1979. "MENT: A Maximum Entropy Algorithm for Reconstructing a Source from Projection Data", Computer Graphics and Image Processing 10, 48-68.

Munson E.S., Eger E.I., and Bowers D.L. 1973. "Effects of Anesthetic-depressed Ventilation and Cardiac Output on Anesthetic Uptake", Anesthesiology 38, 251-259.

Nalcioglu O., Cho Z.H., and Lou R.Y. 1979. "Limited Field of View Reconstruction in Computerized Tomography", IEEE Trans NS-26, 546-551.

Nalcioglu O. and Lou R.Y. 1979. "Post-reconstruction Method for Beam Hardening in Computerised Tomography", Phys Med Biol 24, 330-340.

Papoulis A. 1975. "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation", IEEE Trans CAS-22, 735-742.

Pappenheimer J.R., Comroe J.H., Courand A., Ferguson J.K.W., Filleys G.F., Fowler W.S., Gray J.S., Helmholtz H.F., Otis A.B., Rahn H., and Riley R.L. 1950. "Standardization of Definitions and Symbols in Respiratory Physiology", Fedn Proc 9, 602-605.

Peters T.M. 1973. "Image Reconstruction from Projections", Ph.D. thesis, University of Canterbury, Christchurch, New Zealand.

Peters T.M. 1974. "Spatial Filtering to Improve Transverse Tomography", IEEE Trans BME-21, 214-219.

Peters T.M. and Lewitt R.M. 1977. "Computed Tomography with Fan Beam Geometry", J Comput Assist Tomogr 1, 429-436.

Peters T.M. 1981. "Algorithms for fast Back- and Re-Projection in Computed Tomography", IEEE Trans NS-28, 3641-3647.

PLATO IV Author Group 1974. "PLATO IV Authoring", Int J Man-Machine Studies 6, 445-463.

Robb R.A., Ritman E.L., Gilbert B.K., Kinsey J.H., Harris L.D., and Wood E.H. 1979. "The DSR: A High-speed Three-dimensional X-ray Computed Tomography System for Dynamic Spatial Reconstruction of the Heart and Circulation", IEEE Trans NS-26, 2713-2717.

Rubin H., Geller J., and Hanks J. 1977. "Computer Simulation as a Teaching Tool in Biology - A case study: The Biological Systems Analysis and Simulation Laboratory of the City University of New York", J Comput Based Inst 3, 91-96.

Rushmer R.F. 1970. "Cardiovascular Dynamics", 3rd edn., W.B. Saunders Co., Philadelphia.

Salamonsen R.F. 1976. "A New Approach to Halothane Anaesthesia", M.D.(N.Z.) thesis, University of Otago, Dunedin, New Zealand.

Sato T., Norton S.J., Linzer M., Ikeda O., and Hirama M. 1981. "Tomographic image reconstruction from limited projections using iterative revisions in image and transform spaces", Appl Optics 20, 395-399.

Shepp L.A. and Logan B.F. 1974. "The Fourier Reconstruction of a Head Section", IEEE Trans NS-21, 21-43.

Shim Y.S. and Cho Z.H. 1981. "SVD Pseudoinversion Image Reconstruction", IEEE Trans ASSP-29, 904-909.

Siegel S. 1956. "Non Parametric Statistics for the Behavioral Sciences", McGraw-Hill, New York.

Smith N.T., Zwart A., and Beneken J.E.W. 1972. "Interaction between the Circulatory Effects and the Uptake and Distribution of Halothane", Anesthesiology 37, 47-58.

Smith P.R., Peters T.M., and Bates R.H.T. 1973. "Image reconstruction from finite numbers of projections", J Phys A: Math Nucl Gen 6, 361-382.

Smith S. 1979. "Microprocessor Delivery of PLATO Instructional Material", Presented at 178th ACS National Meeting, Washington, D.C., Sept 9-14, 1979.

Stanton R.E. and Tretiak O.J. 1981. "Dose Reduction in CT Through Variable Exposure Scanning: Computer Simulation", IEEE Trans NS-28, 980-983.

Stark H.H., Woods J.W., Paul I., and Hingorani R. 1981a. "Direct Fourier Reconstruction in Computed Tomography", IEEE Trans ASSP-29, 237-245.

Stark H.H., Cahana D., and Webb H. 1981b. "Restoration of arbitrary finite-energy optical objects from limited spatial and spectral information", J Opt Soc Am 71, 635-642.

Stark H.H., Cahana D., and Habetler G.J. 1981c. "Is it possible to restore an optical object from its low-pass spectrum and its truncated image?", Optics Letters 6, 259-260.

Steward A., Allott P.R., Cowles A.L., and Mapleson W.W. 1973. "Solubility Coefficients for Inhaled Anaesthetics for Water, Oil and Biological Media", Brit J Anaesth 45, 282-293.

Stonestrom J.P. and Macovski A. 1976. "Scatter Considerations in Fan Beam Computerized Tomographic Systems", IEEE Trans NS-23, 1453-1458.

Stonestrom J.P., Alvarez R.E., and Macovski A. 1981. "A Framework for Spectral Artifact Corrections in X-Ray CT", IEEE Trans BME-28, 128-141.

Supan P. 1977. "Feed-back monitoring in anaesthesia. IV: The indirect measurement of arterial pressure and its use for the control of halothane administration", Brit J Anaesth 49, 141-150.

Tam K.C. and Perez-Mendez V. 1981. "Limits to Image Reconstruction from Restricted Angular Input", IEEE Trans NS-28, 179-183.

Tofts P.S. and Gore J.C. 1980. "Some sources of artefact in computed tomography", Phys Med Biol 25, 117-127.

Trussell H.J. and Hunt B.R. 1979. "Improved Methods of Maximum *a Posteriori* Restoration", IEEE Trans C-27, 57-62.

Wagner W. 1979a. "Reconstructions from Restricted Region Scan Data - New Means to Reduce the Patient Dose", IEEE Trans NS-26, 2866-2869.

Wagner W. 1979b. "Incomplete scan geometries in fast X-ray scanners", Biomed Tech 24, 140-147.

Watson G.N. 1966. "A Treatise on the Theory of Bessel Functions", Cambridge Univ. Press.

Wernecke S.J. and D'Addario L.R. 1977. "Maximum Entropy Image Reconstruction", IEEE Trans C-26, 351-364.

Youla D.C. 1978. "Generalised Image Restoration by the Method of Alternating Orthogonal Projections", IEEE Trans CAS-25, 694-702.